HOW MUCH ENVIRONMENT SHOULD ENERGY COST?
(Or How to Find Out)

John B. Collins

November 1974

February 1975 WP-75-20

... What measure is there of the relations of pleasure to pain other than excess and defect, which means that they become greater and smaller, and more and fewer, and differ in degree? For if any one says: "Yes, Socrates, but immediate pleasure differs widely from future pleasure and pain"—to that I should reply: And do they differ in anything but pleasure and pain? There can be no other measure of them. And do you, like a skillful weigher, put into the balance the pleasures and the pains, and their nearness and distance, and weigh them, and then say which outweighs the other. If you weigh pleasures against pleasures, you of course take the more and greater; or if you weigh pains against pains, you take the fewer and the less; or if pleasures against pains, then you choose that course of action in which the painful is exceeded by the pleasant, whether the distant by the near or the near by the distant; and you avoid that course of action in which the pleasant is exceeded by the painful. Would you not admit, my friends, that this is true? ...

...but the behavioural gradient to avoid punishment is always steeper than the gradient to approach reward.

Miller and Dollard

FOREWORD

This I.I.A.S.A. Working Paper has seven essentially simple and straightforward aims:

1) to suggest the essential unity of purpose among alternative strategies which attempt to investigate the Decision-Making Procedure.

2) to offer a framework (the decision 2-space) in which the antecedents, the processing formulas and the consequences of any decision may be seen, critiqued and acted upon.

3) to argue the advantages of an initial 3-pronged excursion this framework; in order to test the comparative effectiveness of these three potential investigation strategies.

4) to explicate how the Direct Scaling technique of S.S. Stenus would approach the Decision Process (and to argue its merits).

5) to argue that the Pandora's Box of problems which seemingly face the decision maker is an illusion: 3 (and perhaps only 2) parametres are sufficient to describe the "value" - past, present, and future - which society places on features of the environment, of energy production and of those social values themselves.

6) to suggest a procedure for generating "events" actual or hypothetical whose consequences and impacts can be examined, scaled along several criteria, and the equivalency of the criteria exactly specified. Progress toward this inter-criterion equivalency is really the heart of this working paper.

7) to offer <u>one</u> experimental route by which these issues might be examined in the context of a study addressing "Public Support for a Nuclear Fuel Reprocessing Plant in the GDR".

For some of these aims, several pages are required to propose, define and rationalize the purpose. For others, often a single diagram, table or paragraph suffices to set the framework for the idea. Implementation is left to the group or person closest to the source of data.

The Paper is intended to be searching and exploratory in tone, rather than rigorous or proof-specific.

TABLE OF CONTENTS

Acknowledgments

Two kinds of contributions merge to enable a Paper of this type to be produced. Both are necessary, neither can be complete without the other.

Surely none of these ideas could have been explored had there been no invitation from I.I.A.S.A. to spend a summer as a research scholar. Time to think, read, ponder and absorb ideas in a nurturant environment was part of I.I.A.S.A.'s contribution, as well as the crass material input to sustain the body: financial and resource support was exceedingly generous; this all through the good offices of H. Swain.

The University of British Columbia provided dual support in allowing me released time to spend the summer in Laxenburg and also in providing office support, "word processing" resources and additional reduced demands during the production phase of this paper.

The second kind of contribution is much more difficult to acknowledge because of its diffuseness, non-specificity, inadvertency or outright unintentionality. These are the contributions from friends and colleagues in form of ideas, communication, arguments, lent books, reference, bull sessions, etc. How these scattered kernels of thought coalesce into a unified fabric must remain a mystery. Persons that are spcifically due thanks however are: Harry Swain, Wolf Haefele, Greg Baecher, Howard Raiffa, Amos Tversky, Earl Cook, James Coleman, Richard Meier, Tjalling Koopmans, Valery Sokolov, James Miller and Charles Wolfe. In each person's case, some comment or idea during conversation found its way into these notes. The conversational sessions ranged from day-long ordeals with Greg Baecher to a brief lunch time chat

with Professor Koopmans. From Baecher came the notion to test perceived likelihood vs. perceived impact, given the event, as separate issues, on the bet that one of the features which distinguish between environmentalists and the "development mentality" in assessing risk are differences in perceived likelihood, not impact, of a hypothetical event. Professor Koopmans inadvertantly used the word "doomsday" during lunch one day, in talking about energy pricing, prompting me to read his Paper "Proof for a Case where Discounting Advances the Doomsday". From that experience emerged the idea to test the "discount rate" (in curious usage) of perceived risk in the past, future and in space.

The thrust to develop and include the space/time graphs derived from two raised eyebrows and a query from Wolf Haefele, whether that was my own idea.

It would be tedious to comment on each contributor in that list, but thus these ideas weave themselves into an eventual framework.

A particular point of thanks is due to those forces at I.I.A.S.A. for forbearance in not foreclosing, thus allowing these ideas to incubate and deliver themselves.

INTRODUCTION

The decisions which now face the world are momentous and unprecedented in their magnitude. For the first time in the world's history the human race must decide issues and develop plans and strategies which will affect the shape of the world to come, and with whose consequences we must live for decades if not centuries.

Thus I.I.A.S.A. and similar institutions are charged to anticipate, foresee, and develop these decisions as well as routes, costs and benefits of accomplishing them.

Their multi-national, and worldwide character presupposes that aggregates of nations must be responsible in this planning. While all nations face approximately common problems, still internal policies, differing needs, and competing philosophies demand that several routes be examined in order to accomplish the respective goals. Surely there is no ideal decision strategy known yet.

The implication of the decisions now facing us are overwhelming, particularly since many could not have been attempted before; hence, there is no storehouse of experience from which to examine the consequences of both positive and negative outcomes of the decision. Conceivably, one single such negative outcome could eliminate much, if not all, human life in an entire region. These are decisions of the "what if" variety. "What if an explosion were to occur?" "What if large quantities of nuclear material were to be diverted and used in international blackmail?" These consequences lie in the realm of the

"hypothetical" and are more fully explored in a recent MINERVA article by Haefele. Central to that article and this Working Paper is the concept that "what if" questions can only be assessed by examining them in the fuller context of problems whose consequences are known.

This Paper explores one systematic method for obtaining information about public assessments of such hypothetical questions, and indeed does so by scaling "what if" questions simultaneously with other questions with known consequences. Provisions are made for three concurrent approaches to bring the merits of each to bear on that same general question. Pareto-optimal and Multi-objective personal assessments of the differential consequences of several potential decisions. In general, the guiding thought is that Magnitude Estimation Techniques will form the "broad brush-stroke" context of many peoples' assessments of a great range of events -- simpler and without necessity of prior training. Subsequently, Pareto-optimal and Multi-objective methods should be used to "fine-tune" or improve the "resolution" of peoples' assessments. Concurrently, potential differences among special interest groups representing divergent platforms will be tested by each of the three assessment techniques to uncover 1) differences in perceptions of impact of an "Event" on some criterion, 2) different weightings among criteria or 3) additional descriptions of group membership.

A broadly focused Task List outlines the general way to accomplish these goals.

Some Definitions: As is sadly, but inevitably the case, the jargon of one discipline is rarely the same as that of another. Thus, to specify early, how certain terms are to be understood has considerable advantage.

Simply put, there are three central terms, surrounded by several other supporting concepts. These most often mentioned terms are Events, Criteria and Persons (groups). A brief explanation of each, and their interrelationship now will save much confusion in later reading.

Event - the occurence, given or hypothetical, of a situation, happening, or outcome where the occurance bears interest to I.I.A.S.A. because of its impact on the environment, social fabric or economic sector of a city, region, etc. Events of interest may (and indeed should) range in impact from the near trivial to the potentially cataclysmic, in order to assess their positions, in context of each other, on one or more criteria.

Criterion - dimension, explicit or implicit, surface or latent, on which relevant events can be ranged. One might"assess the relative advantages of several nuclear power plant safeguard systems"on the criterion of Cost, or of Risk Protection, or of Present Availability -- where Cost, Risk Protection and Availability are the Criteria. Clearly, some events are not relevant to certain criteria; further, relevant events assessed high on one criterion may justifiably be assessed low on others. The central issues of this Working Paper are twofold: 1) how to achieve explicit measurements on criteria of subjective values such as "Degradation of the Visual Environment" on "Publicly Perceived Risk Reduction"; and 2) how to demonstrate the equivalencies or trade-offs which obtain among all measurable criteria, both subjective and "hard", (cost, energy requirements, etc.).

Person (Group) - the individuals who assess events on various criteria. Obviously such individuals will often represent vested interests, organized platforms or extreme positions. Where demonstrated, these persons can be deemed spokesmen (formally or otherwise) of various special interest groups. A central question to examine is whether these groups have any uniquenesses which cannot be specified as a particular assessment of an event or a criterion, and a particular weighting of that criterion vis a vis other criteria.[My own hunch is that group membership (when it exists at all) can be fully specified by these two parameters.]

Assessment - evaluating (preferably in quantative terms) the relative merits of a series of events on an appropriate criterion. Assessment procedures might include: priority ranking, equal internal scaling, direct scaling (magnitude estimation), indifference tradeoffs, etc. Obviously, the more clearly the assessments can reflect the most powerful atttributes of the number scale into which they are translated, the greater utility the assessment has. Therefore, one would lean away from priority ranks and toward magnitude estimates, on probability function, both of which can be translated into Dollar equivalents.

Embeddedness Context - the property possessed by events ordered on a criterion, that the impact of one event (in terms of that criterion) is far better understood when examined in the context of other events ordered along that same criterion. Haefele's MINERVA article exemplifies that the impact of an event can only be understood insofar as it is embedded in the series of other events where impacts are similarly scaled. This working paper takes that notion two additional steps: 1) that the contextual ordering surely varies

from criterion to criterion, but can be measured and, 2) that it is the trade off function among the criteria that are of greatest interest; the specification of an event in its context on a single criterion is but a step en route to the trade off functions.

Thus, the unitary concept is then an event may be assessed on a criterion by a person. Or, in expanded form , many events assessed on a criterion will demonstrate a particular contextual configuration -- some assessed high on the criterion, others low. The same, identical set of events will exhibit a different ordering (contextual configuration) on a second, but appropriate criterion. Finally, different individuals will demonstrate differing opinions about this ordering of events; hence, will show differing personal expressions of the contextual configuration. It is yet unproven whether persons who represent a particular special interest group will have essentially similar configuration patterns, as well as similar weightings of the criteria.

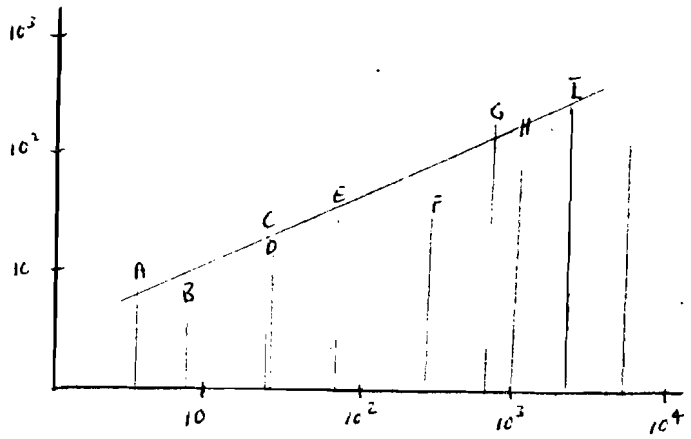


Figure I.

An instance where an indeterminate number of persons have assessed 10 events on each of two criteria. The impact of any single event on one criterion can be determined by examining its projection onto that criterion, thus showing the contextuality or embeddedness of that event amont many others. The relationship between the two criteria is given by the slope of the line. The choice of log-log coordinates will be explained later.

2. The Geometry of the Decision Space.

It is a curious enigma of decision-making that while the consequences of that decision endure in space and time, the nature of the decision-making itself, its origin, its deliberation, and its rationalization all exist in a separate realm of space/time not normally made explicit.

Examine the 2-space in Figure 2. The axes are Time and Space, where the time axis extends from the distant past on the left (10^4 years ago)through the present and into the far future on the right (10^4 years hence). The space axis extends from remote space at the bottom (10^8 meters away) toward oneself

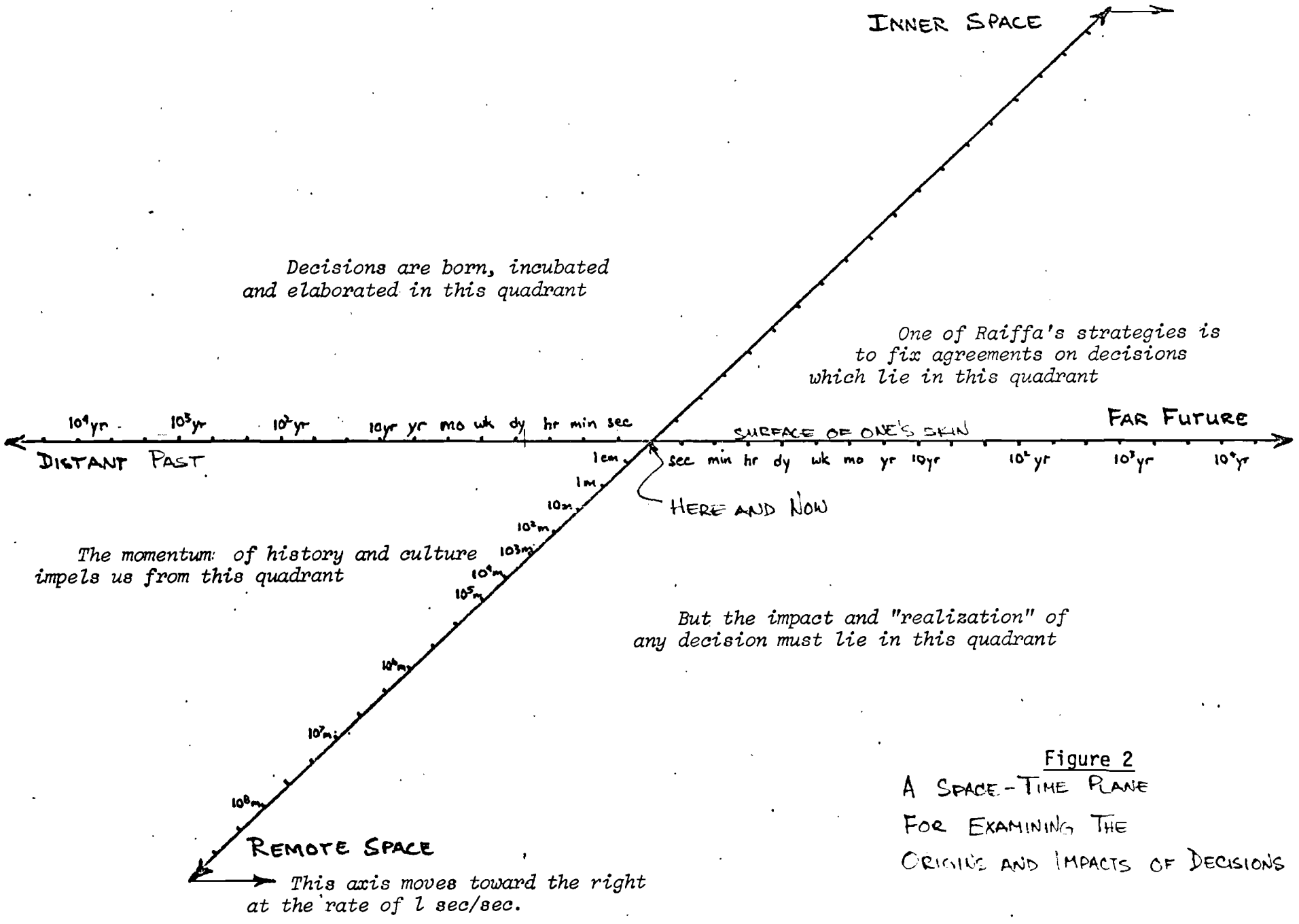


Figure 2

A SPACE-TIME PLANE FOR EXAMINING THE ORIGINS AND IMPACTS OF DECISIONS

up to one's skin surface at the origin ; but then proceeding inward toward a conceptual "inner space" not yet well investigated. The units on the time dimension of Figure 2 are to be taken as conceptually useful, but non-mensurational in that the scale is sharply stretched toward the present -- an advantage in detailing events proximate in time. Similarly toward the bottom of the figure, the units, in meters, are sketched. The remaining dimension "units of inner space" is perhaps best left suggested, but undefined, although a later Figure will indicate some landmark units. These units are likely dependent on one's intellectual, personal, moral, and religious convictions.

The decision paradox is, of course, that it is precisely in this "inner space" that most decision-making effort is expended in conceiving, elaborating and actualizing a decision.

For any decision to achieve reality in tangible space-time, that boundary must be crossed which lies at the origin -- the here and now. Otherwise, an embryo decision remains an interesting notion, a conceptual artifact, but not a tangible reality. Note, however, that the substance of the idea of the decision can be transmitted indefinitely forward in time by communicating the information to others, but without its ever crossing the skin surface barrier. (An interesting diversion is to speculate whether the energy required to cross the skin barrier is equivalent to the energy needed to produce a localized reversal in the Second Law? Can Shannon thus be linked to Relativity?

A moment's examination of Quadrants II and IV will show their proper perspective in contributing to the summation of forces leading to Decision Implimentation.

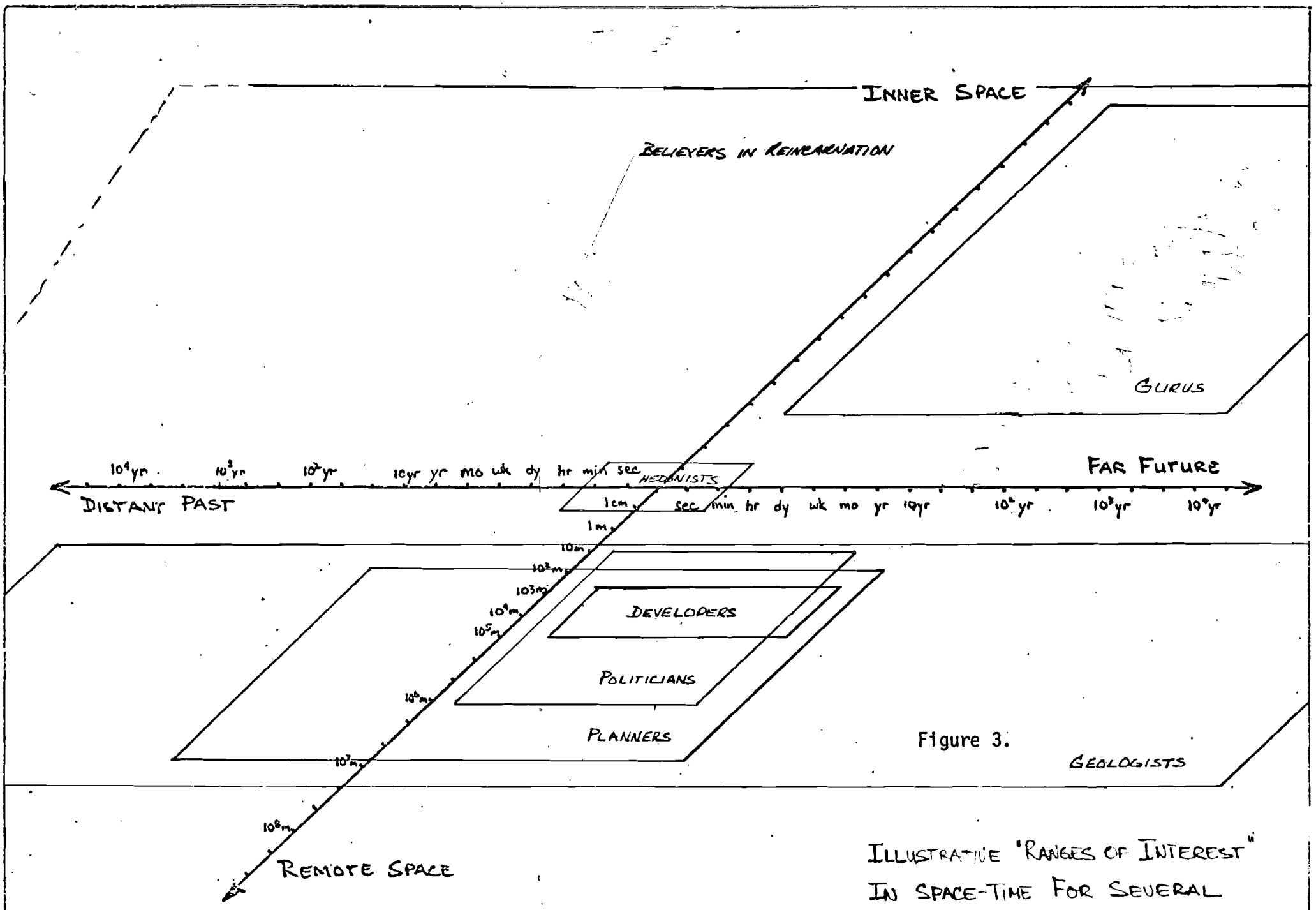


Figure 3.

ILLUSTRATIVE "RANGES OF INTEREST" IN SPACE-TIME FOR SEVERAL PROFESSIONAL GROUPS

Extending this strategy then, it is helpful to elaborate the "ranges of interest" of several professional groups. Figure 3. documents one conceptualization of eight such "ranges of interests". Beginning nearest the origin, one might speculate (perhaps unfairly) where interest domains lie for the hedonist-from the ecstasies of a moment ago to the anticipations a moment from now. Further, the interests of the politician may extend about ten metres beyond his immediate person up to 10^6 metres or 1,000 kilometres if he is a very powerful politician. Along the time dimension his interests may range from a few second hence the lapse of his mandate however long that might be.

Land and property deveopers, may have restricted space ranges at their disposals, a hundred metres up to ten kilometres; and a time range from the next several minutes to some ten year; provided that the developer's chief interest is to assemble the land, construct on it, and sell it off. Planners, contrariwise, have ranges of interest extending well into the past for those who are historically aware planners. Range of spatial interests may encompass from a hundred metres to perhaps nationwide dimension 10^7 metres; temporarily from a hundred years ago to the same point in the future. The interest range of geologists tax the limits of the graph since they extend into the far distant past and forward into the far distant future. (The age and lifespan of the earth are probably not proper topics for this paper.) His spatial domain may span those soil samples at his very feet remotewards to the opposite side of the globe, some 1.2×10^7 metres away

"Professional" interests in other quadrants are harder to describe. The interests of the guru perhaps spans all of inner space and far into the future. And certain schools deny the need for consideration of any factors other than the immediate

here and now. Religious leaders advance teachings pertinent to this quadrant.

Quadrant III, the inner, past is more difficult yet. Believers in reincarnation and transcendentalists would fall into this quadrant, although probably not exclusively.

An exceedingly comprehensive range of interest belongs to the astronomer. Sketch it for yourself, if you choose.

One handy feature of this decision 2-space is that additional dimensions may be freely added. In the third dimension one can easily represent criterion such as:

number of lives affected by a decision.

time required to achieve a program.

cost (dollars or man years) to accomplish a a change in policy or practice.

"perceived magnitude" of risk

Each criterion represented as a third dimension, yields a surface in space/time and (for instance) perceived risk. One then obtains a clearer understanding about the nature of public willingness to accept change or the consequences of such change, or even of the costs of implementing some program. More helpfully, one sees for oneself the extent in space, time and impact, what the likely consequences of any particular act, policy, program or risk will probably be.

The poor power of geometry cannot show the multidimensional case, where 4,5, and 6 criteria are all analyzed simultaneiously, but the mathematical techniques to do so are commonly available and can be implemented on demand. A later section will address itself to which criteria? how many? all at once? how

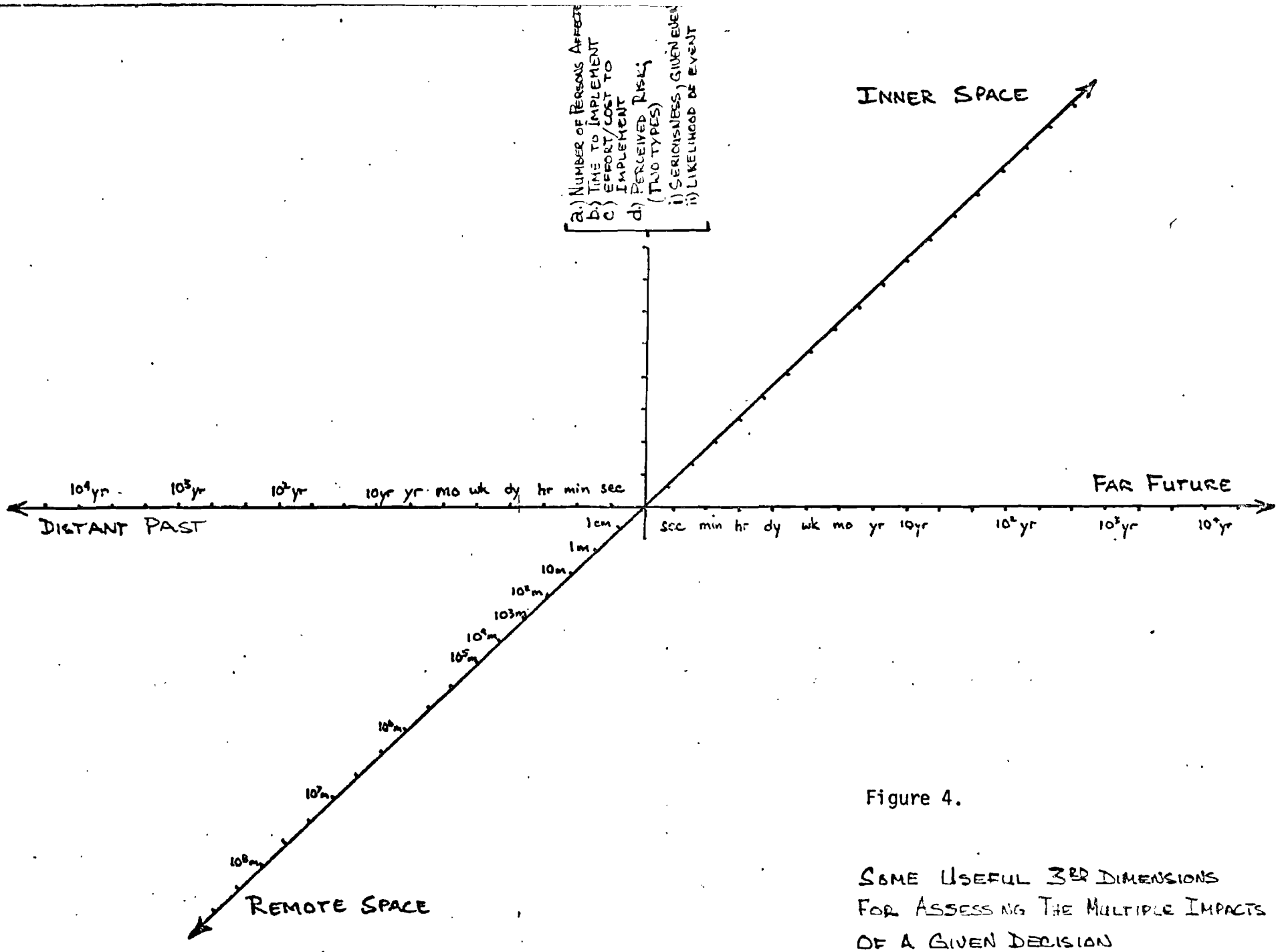


Figure 4.

SOME USEFUL 3RD DIMENSIONS FOR ASSESSING THE MULTIPLE IMPACTS OF A GIVEN DECISION

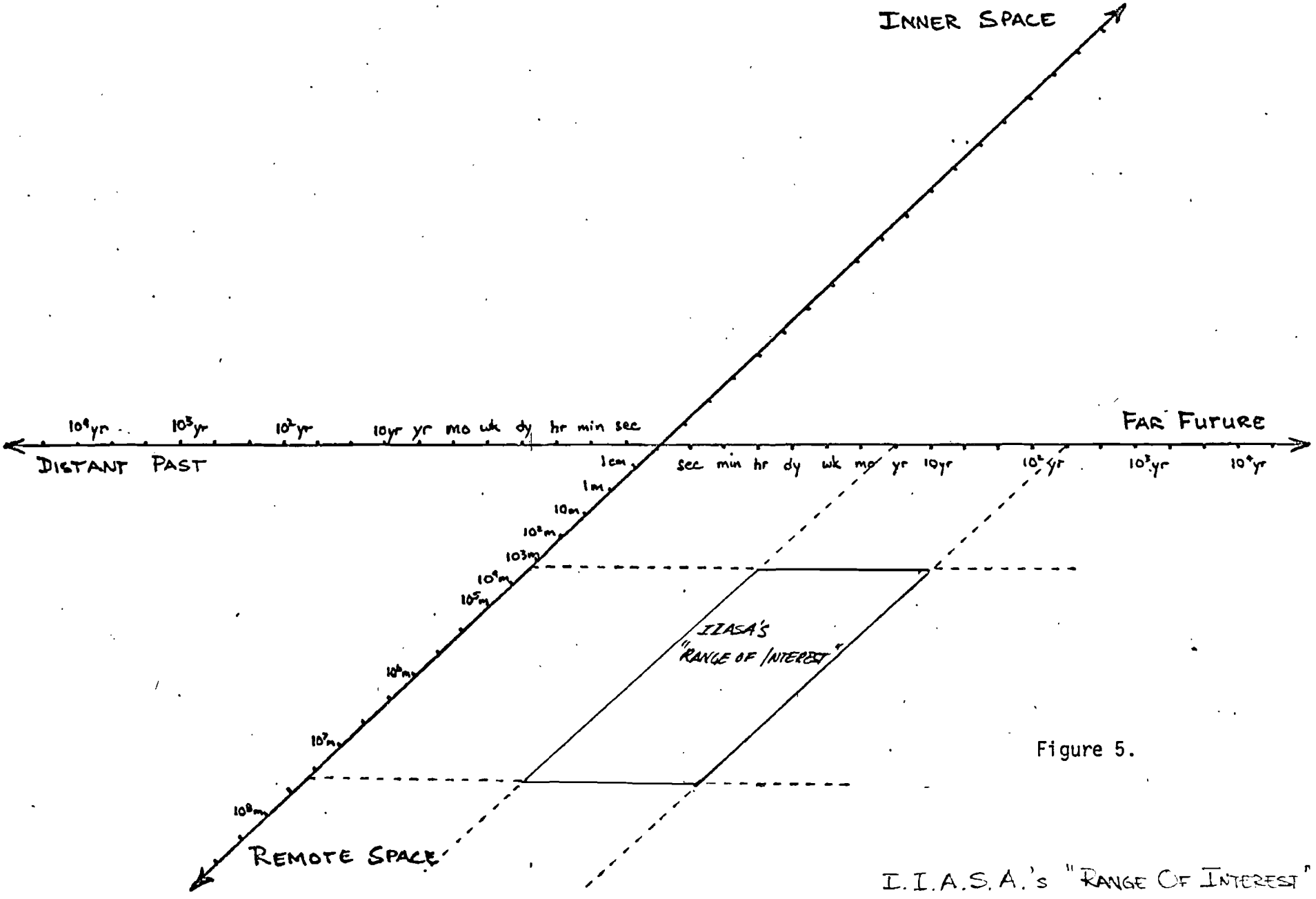


Figure 5.

I.I.A.S.A.'s "RANGE OF INTEREST"

many people? how many groups? what kinds of events?

With this convenient space/time tool, it may be instructive to focus it inward for a moment. I.I.A.S.A. itself, eclectic and polymorphous as it is, nevertheless has some identifiable if approximate boundaries. Figure5 illustrates the outermost boundaries which should contain all ranges of interest of the I.I.A.S.A. subgroups. The spatial span may be deemed to range from the single neighborhood upwards to encompass the entire planet's surface. Temporally most project's impacts will begin in a year or so, but may have consequences (or at least examine problems) for the next 100 years.

That the ranges of interest so conspicuously exclude the past may give some of I.I.A.S.A.'s members cause to wonder whether there is merit in considering projects which more completely span the "temporal landscape" as well as the spatial.

At a yet closer focus (and adding a new dimension), I would assess the approximate location of the Urban and Regional System project as shown in Figure 6. The interest domain of the Project (for some of its members) may be so sharply focused as to examine individual housing units (10^2m) while for others, regional systems extending through hundreds (or even thousands) of kilometres may be the unit of analysis. Temporally, the earliest one can reasonably hope for substantial data is about a year from now, but if the project grows as intended, urban and regional management for the next century becomes the chief focus. But how many people does this affect? At the moment, not many. U.R.S.'s impact may be largely restricted to the I.I.A.S.A. staff members who know about it, and a handful of others who have corresponded with the Project. In future, though, the management strategies investiaged by U.R.S. may come to affect

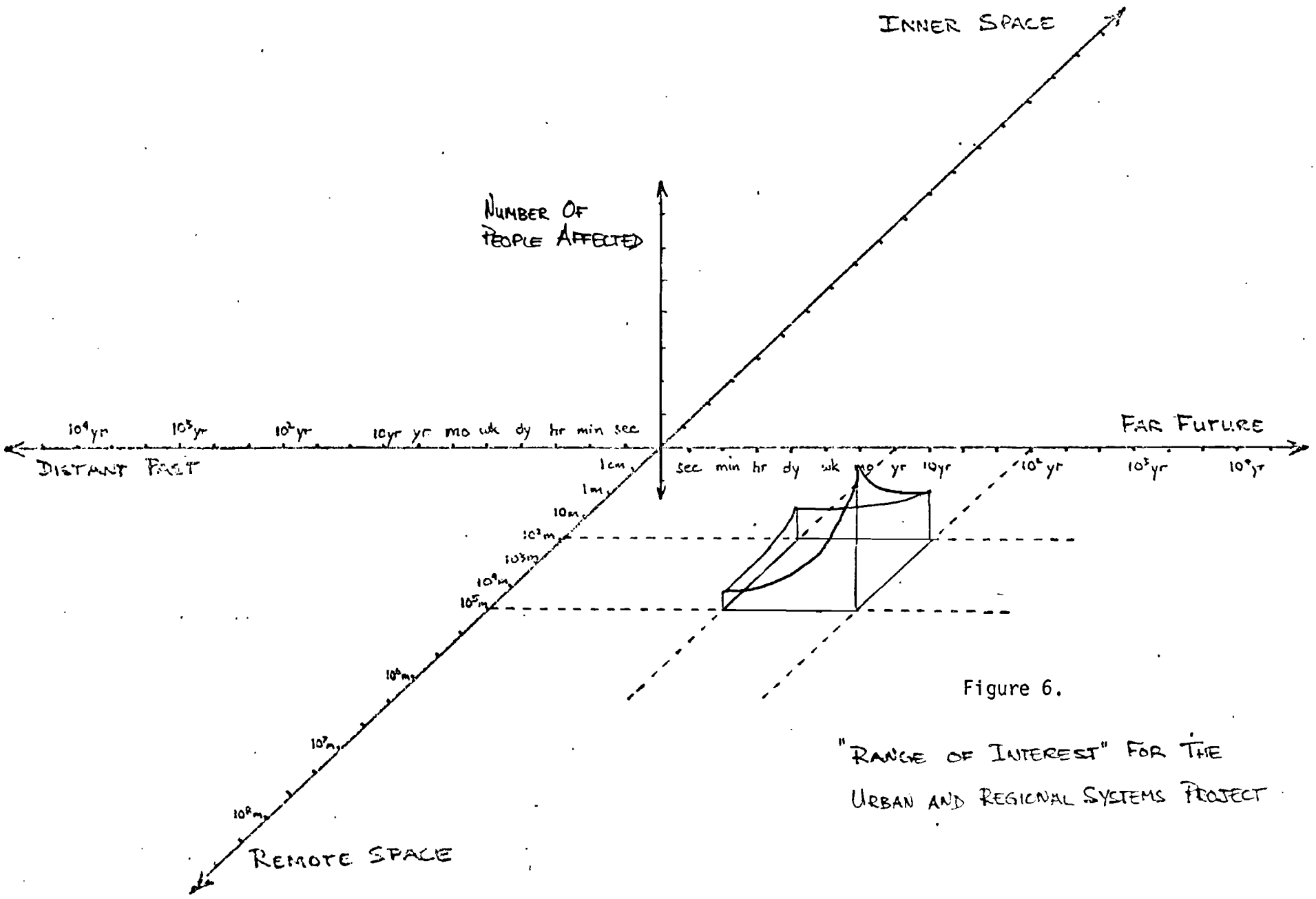
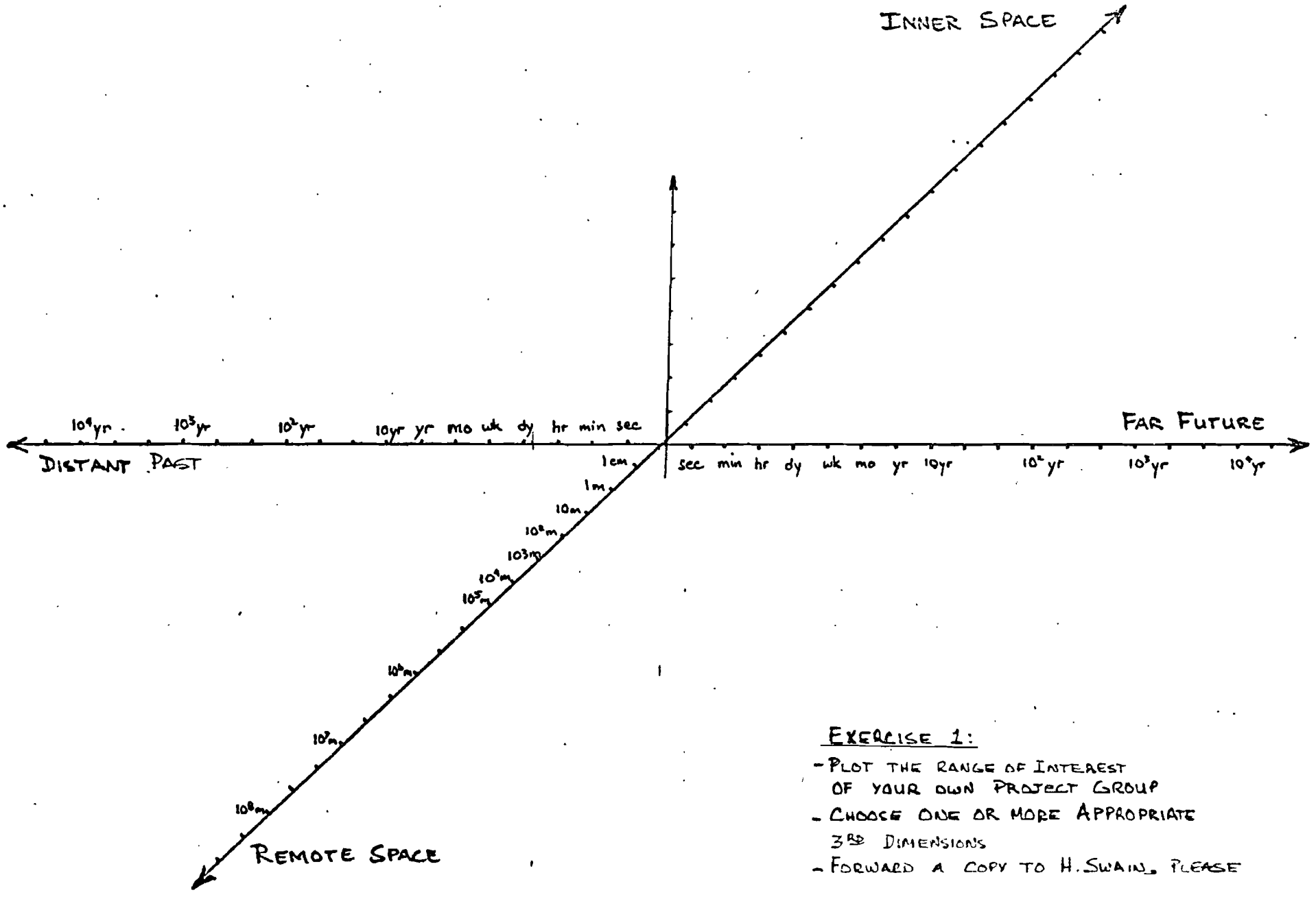


Figure 6.

"RANGE OF INTEREST" FOR THE URBAN AND REGIONAL SYSTEMS PROJECT

(in quite direct ways) tens of thousands if not millions of people. Thus the figure cants sharply upward for future, and more cumulatively distant loci.

The reasons prompt the Exercise on the following: 1) to walk through the time/space location of ones own Project delineates rather clearly what one purports to be all about (self interest) and 2) it should be or more than passing interest to plot each of the Projects and sub-groups on a common 2-space to examine the temporal and spatial overlap, but see the 3rd dimension diversity. Thus I (at my 10^5 metre removal) have designated Harry Swain as the common collection point. He will collate the materials he receives.



EXERCISE 1:

- PLOT THE RANGE OF INTEREST OF YOUR OWN PROJECT GROUP
- CHOOSE ONE OR MORE APPROPRIATE 3RD DIMENSIONS
- FORWARD A COPY TO H. SWAIN, PLEASE

Two Principle Structures.

A complete exploration of the decision-making domain must explore the space/time correlates of the two structures most centrally involved; the Self which generates, elaborates and implements a decision, and the World which receives, houses, and concretizes the results of that decision.

The World forms the storehouse of all past decisions which have broken the "skin-barrier" and have been actualized in object form. As such, the vast momentum of culture, history, more and tradition, lies to the left of the time axis. This residual collection of all past decisions profoundly affects the infinity of potential futures which lie to the right of the time axis since of all conveivable futures, the most likely one is the straight line extrapolation of today, driven by the momentum of the past.

Should one choose to assign a probability value to each of the possible futures, the highest value must go to this extrapolated present. But if we examine that most probable of futures and decide that however probable, it is somehow undesirable, then the only way the profile of probability for all possible futures can be altered in application of forces directed through the Self to realign the probability profile. Indeed, merely thinking about an alternative future increases its likelihood above the background probability. To take specific steps to bring it about raises its likelihood yet higher. But in all such cases, historical momentum (outside the skin surface) is redirected and focused (within the skin surface) to cause a different future to come about (once more outside the skin surface).

The effects which are thus brought about may be trivial -- i.e. extending only briefly in time, and not far away from one's person; or they may be monumental in time and space.

The forward two quadrants in Figure 7 illustrate in Ziggurat - fashion the stability in time of certain substructures in that external World. Nearest one's skin surface -- a scant meter in diameter -- is the Personal Bubble (detailed by Hall) which lasts only a few second as one moves through space. One's territory, home, garden, etc., may extend some 10 meters on either side of one's person, but persist notably longer -- as much as a few decades, even. A next higher level of aggregation will comprise one's neighborhood -- perhaps 100-200 meters to either side spatially, and another few decades temporally. Cities and Regions may be gauged similar in temporal extent (80-200 years) while accreting space an order of magnitude additional each. Nations vary in size, but 1,000 kilometers either side of one's person will include the boundaries of most nations. Temporally 100-200 years serves as a tolerable "average national lifetime". There can only be two longer aggregations -- continents and planets at about 5×10^6 and 1.2×10^7 meters respectively -- but whose temporal extents are measured in units of Geological Time.

The Self can only be diagrammatically rendered due to the profound uncertainty of the spatial units, even though the signposts of a man's lifetime can be clearly specified on the time dimension.

In arbitrary units, though, the Self can be conceived (NPI) in five Zones. Proceeding inward, the flesh, skeleton and material substances of the Body comprise the first zone. Memory, the storehouse of personal experience and learning forms the second. To this should be added the vicarious memories of

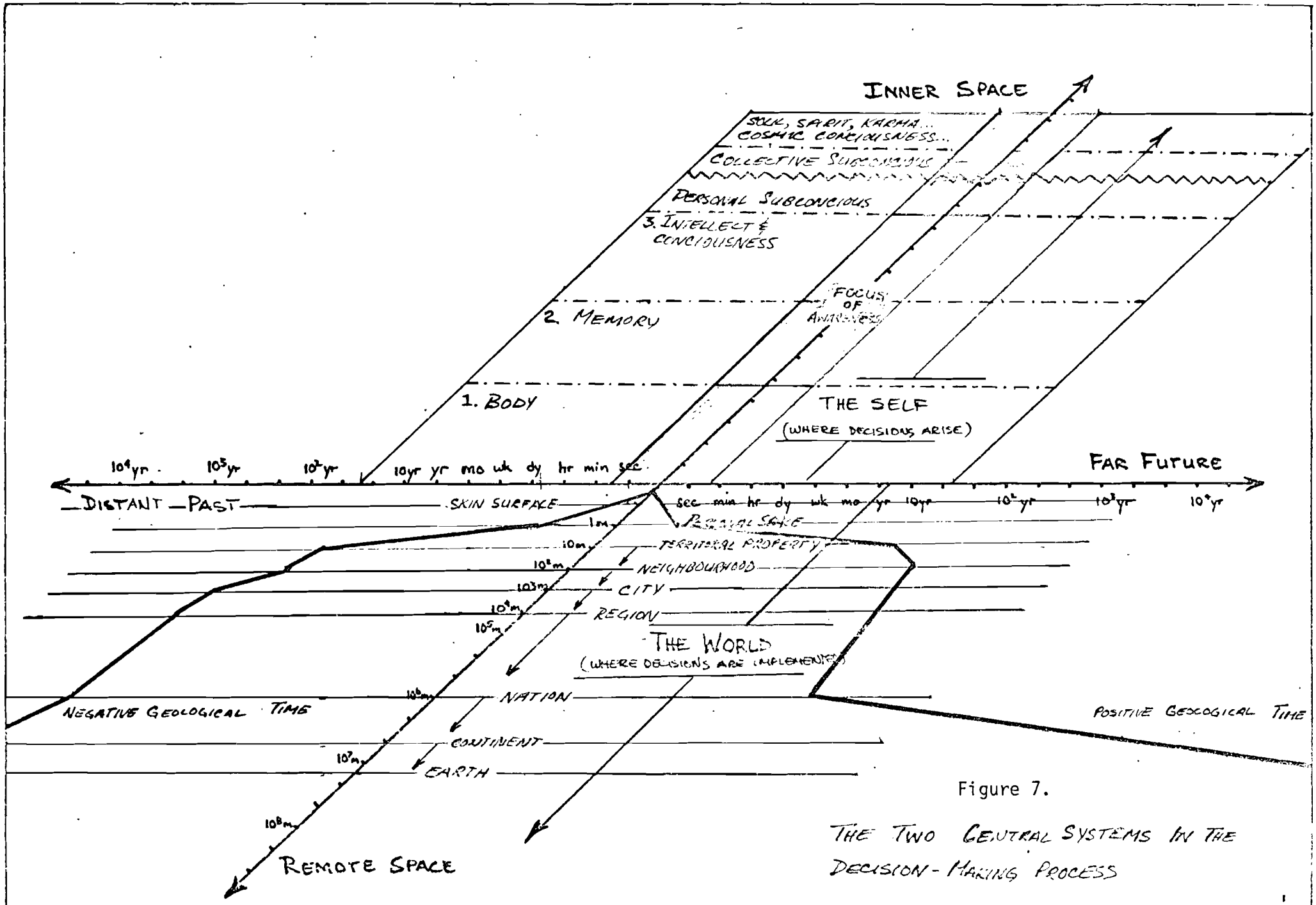


Figure 7.

THE TWO CENTRAL SYSTEMS IN THE DECISION-MAKING PROCESS

friends, teachers, books read, knowledge gleaned from second and third party sources. The third Zone can be construed as a Central Processing Unit of the Self -- its consciousness, intellect, extant problem - solving and decision-making styles, that part of the Self which is directly and immediately aware. Zone 4 displays a dual character (and is profoundly Jungian) in that it comprises the Personal Unconscious at its outer level and Jung's Collective Unconscious at its inner. The final and innermost zone (5) is surely the most mystical and therefore the most arguable. Characterize it as Soul, Spirit, Cosmic Oneness -- as you will. Perhaps it need not be invoked for this analysis, but it should also be neither forgotten nor denied.

Orthogonal to these Zones (notably) and centered on the Present time axis is the focus of awareness of the Self. The diagram shows it as a discrete zone; probably it is not. Rather it is better conceived as shading diffusely forward and backward in time, and clearest in the immediate present. Accomplished Masters of Eastern Philosophies maintain that as we become Enlightened, the focus of Awareness at the immediate present subsumes all things past and future. Until we achieve that happy state, the Figure is a convenient heuristic to direct our attention toward experiential sectors we might otherwise miss.

FLOWCHARTING A DECISION

Whatever the ontogenetic "primal cause" of a decision, the post-partum evolution is quite easy to trace. Consider Figure 8. In deference to that Nietzchean battle between Appollonius and Dionysius, let us assume that the "Aha!" experience arises in Zone 4 at the personal/collective frontier. (Platt alludes to this as the discrepancy of "what is vs. what might be"). In this Figure, the breaking awareness at A that a decision needs to be made is transmitted instantly forward to the CPU at B. The CPU directs a search of the relevant

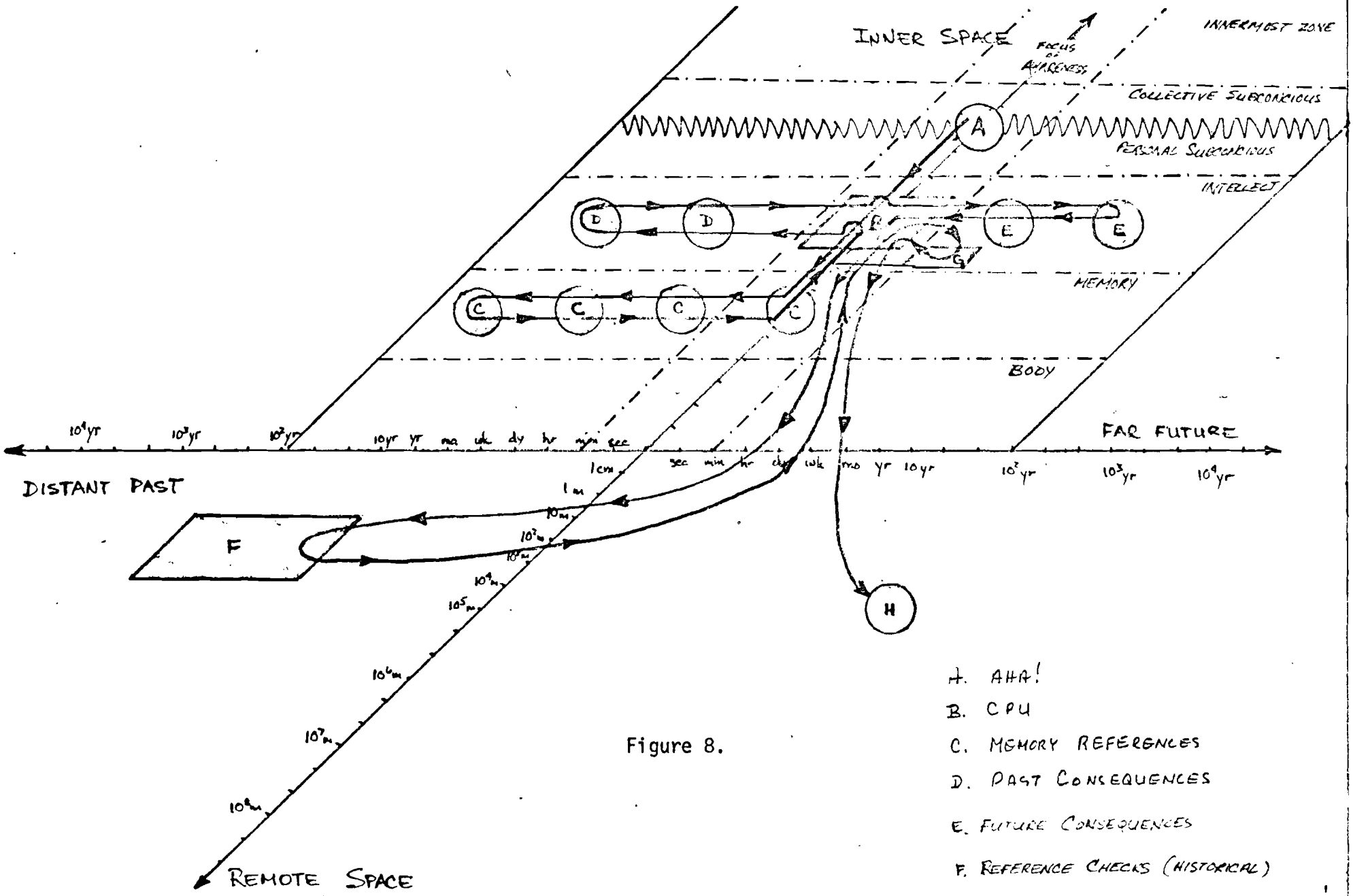


Figure 8.

A. AHA!
B. CPU
C. MEMORY REFERENCES
D. PAST CONSEQUENCES
E. FUTURE CONSEQUENCES
F. REFERENCE CHECKS (HISTORICAL)
G. DECISION MAKING STRATEGIES
H. IMPLEMENTATION

memory storage C. With this additional information the CPU directs a historical "what if" strategem. "What if this situation were already the case, what would have been the results?" (Hypotheticality in the past). The CPU then directs the same strategem forward into the future, E. "If this situation were to transpire, what would its outcome be?" (Hypotheticality in the future). Thus for the entire decision processing may have taken no more than a few seconds. So armed with memory and two kinds of hypotheticality, the CPU may then order a search of events in real space and past or present time. To seek advice, search libraries, reference experts all require that the CPU obtain inputs from outside the skin of the Self, collectively shown at F. CPU next implements the most complex of all processes -- what to do. The cumulative inputs from C,D,E and F are all directed toward G for processing according to the prevalent Decision-Making Strategies. By far, the largest fraction of I.I.A.S.A.'s efforts are directed toward this single component of the sequence -- how to decide. (Perhaps the next largest fraction derives from A -- what to do). Most certainly, the primary reason for this Working Paper and its eventual incorporation in a Test Case is to illuminate that same G component. If these G Strategies return a "go" decision, the CPU must then marshall its forces to enact H -- Implementation -- the only other time in the entire sequence that involves crossing the skin barrier.

Why dissect the decision sequence in this detail? Two reasons: First, it highlights the "embeddedness" of the decision-making component within the sequence. While it is a vital component, it is still only a part of an entire system, dependent on the operating characteristics of seven other components and upon the quality of the information flow within the system's communication network. GIGO is no less true here than in any other information processing

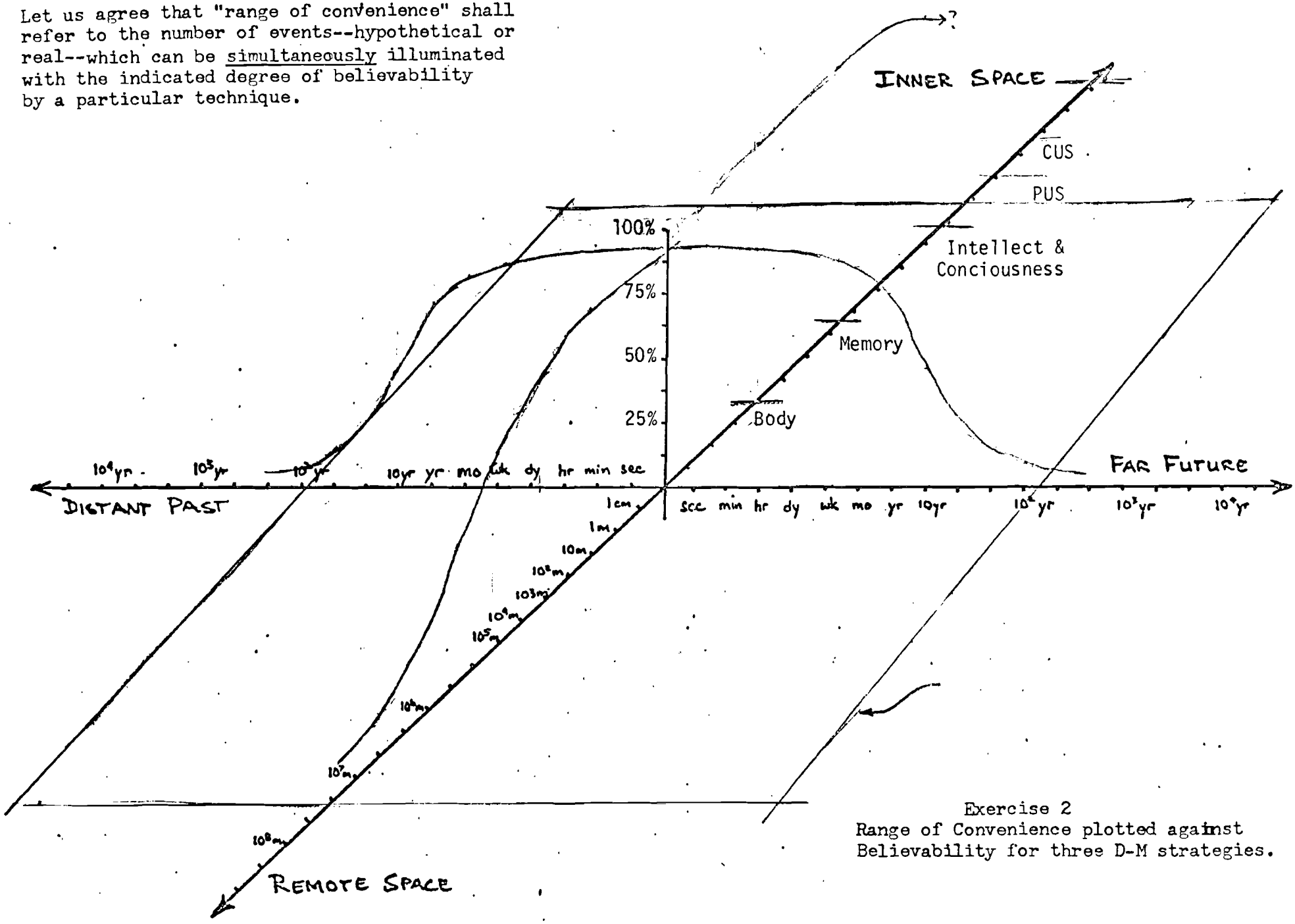
system. Secondly, it sheds some critical wavelengths on the Hypotheticality issue. Hypotheticality is thus shown to be a feature peculiar to Inner Space; indeed restricted to Zone 3 of Inner Space. Hypotheticality does not exist in External Space in either past or future time, but can exist in both past and future quadrants in Inner Space.

But not only is Hypotheticality an artifact of Zone 3 Inner Space, so are all the embedded elements (events) in whose contexts the hypothetical events are to be examined. In short, while Haefele cautions that NO probabilities associated with Hypothetical occurrences made zero; so also can no probabilities (however desirable) be made unity -- the very deliberation of these contextual probabilities is forever a Zone 3 phenomenon. It is how we feel about the probabilities, not what the probabilities actually are, that drives us to ever more explicit decision techniques.

Bear in mind then, that all 3 decision tactics under scrutiny here are Inner Space tactics. They may "game" with events which have happened (Real Space, Past Time), which might happen (Inner Space, Future Time), which might have happened (Inner Space, Past Time). Some events are non-hypothetical, but the gaming always is. (Note also, that the remaining quadrant, Real Space, Future Time can only be examined via Inner Space, Future Time.)

A useful, if soul searching, experience will induce the proponents of each of the three Decision-Making strategies to bound the "range of convenience" of their pet technique in this decision 2-space. The results for Stevens' Magnitude Estimation follow.

Let us agree that "range of convenience" shall refer to the number of events--hypothetical or real--which can be <u>simultaneously</u> illuminated with the indicated degree of believability by a particular technique.



Exercise 2
Range of Convenience plotted against Believability for three D-M strategies.

The Merits of Magnitude Estimation

The choice of Steven's method of magnitude estimation as the most utilitarian route to the measurement of public values presupposes but a single assumption; indeed, a rather non-psychological one: that the best way to procede is to ignore all mediating variables (attitudes, opinions, socio-economic status) and to procede directly to the issue in question - - Public Values. That this is probably the correct route to follow is based on several explicit reasons: 1) Experience in the Vancouver Urban Futures Project shows that there is a significant, but low correlation between analogous issues in the Urban domain and such mediating variables. It appears doubtful that the correlations have sufficient predictive validity (whatever their significance level) to make them useful tools. 2) The usual applications of mediating variables is to assist in explaining some previously measured target behaviour; but for IIASA, the Social Values themselves are the target behaviours and thus should be approached directly. 3) The target behaviours (Values) are more explicit, hence more measureable than the attitudes which may underlie them. 4) Attitudinal and dispositional information is certainly interesting, but is a secondary importance to IIASA at the moment, therefore should be delayed until the primary purposes of this Test Case are accomplished.

Just how uninteresting the relationship between personal dispositions and perception of Urban problems really is can be seen at a glance in Table 1. Priority Rankings for 29 Urban problems were used as predictors of McKechnie's 9 scales from the Environmental Response Inventory in a stepwise regression analysis. Cumulative predictive

powers of the 29 predictors are indicated for each of the 9 ERI scales.

Table 1

Predictive Power (R^2) of 29 Ranked Urban Problems Against 9 Scales of the Environmental Response Inventory

Pastoralism	.19
Urbanism	.06
Environmental Adaptation	.21
Stimulus Seeking	.14
Environmental Trust	.17
Antiquariansim	.10
Need for Privacy	.03
Mechanical Orientation	.04
Communality	.06

Table 2 shows similar entries when traditional Socio-economic measures were used to predict the same 9 ERI scales. Predictors for this case included sex, age, education, income, property taxes, rent, time and mileage from the workplace.

Table 2

Predictive Power (R^2) of 12 Socio-Economic Indicators Against ERI Scales

Pastoralism	.11
Urbanism	.08
Environmental Adaptation	.08
Stimulus Seeking	.22
Environmental Trust	.22
Antiquarianism	.16
Need for Privacy	.07
Mechanical Orientation	.26
Communality	.05

Clearly, immediate attention should focus on the target variables themselves - - their purification and quantification, and only later on associated variables, however interesting, which may help explain the "Why's" of public values.

How Magnitude Estimation Operates

Because not everyone will be aware of the Magnitude Estimation Technique, much less of its application to such seemingly "soft" variables as Risks, Environmental Esthetics, Energy Benefits, or others of central interest to this Test Case, what follows is a rather lengthy explication of the tecnique; its prior validation, merits and hazards. The passages are excised directly (very directly) from Stevens' article in Science, February 1966, A Metric for the Social Consensus. I have injected the occasional question, emphasis or comment from other sources; they are immediately recognizable.

A Metric for the Social Consensus

S. S. Stevens

Methods of sensory psychophysics have been used to gauge the intensity of opinions and attitudes.

Experiments in a dozen laboratories have shown how procedures developed for the scaling of sensory attributes such as brightness and loudness can measure human reactions to many forms of nonmetric stimuli. The procedure called magnitude estimation, for example, has been used to gauge the consensus concerning intensity or degree for such variables as strength of expressed attitudes, pleasantness of musical selections, seriousness of crimes, and other subjective dimensions for which the stimuli can be arrayed only on nonmetric or nominal scales. These applications of direct scaling presage an advance beyond the indirect methods developed by Thurstone (*1*), who first brought the logic of Fechner's psychophysics into the domain of attitudes and opinions.

If, as many have believed, progress in the behavioral sciences demands the creation of tools to quantify such elusive variables as opinions, attitudes, preferences, esthetic values, utility, and so forth, the outcome of more than a dozen different experiments lays a foundation for optimism. On many attitudinal continua, the

It has been learned that observers can match numbers to stimuli and stimuli to numbers; they can estimate the apparent ratios between stimuli, and they can adjust stimuli to produce prescribed apparent ratios. All these methods give results that are related to the stimulus values by a power function. The power function is a necessary consequence of the ratio invariance of the psychophysical law. If equal stimulus ratios produce equal perceptual ratios, the perceived magnitude ψ grows as the physical value φ raised to a power β

$$\psi = k\varphi^{\beta}$$

The measure of φ begins at threshold; k is a constant that depends on the units used. Each modality or continuum appears to have its characteristic exponent, ranging in value from 0.33 for the brightness of luminous fields to about 3.5 for the apparent intensity of electric current passed through the fingers. A convenient feature of the power function is that, in log-log coordinates, it takes the form of a straight line. The slope of the line gives the exponent of the power function.

Perhaps the most reassuring development in psychophysics is the validation of the various exponents by the procedure of direct cross-modality matching. Just as lights of different hue may be matched for brightness (as in heterochromatic photometry), so sensations in one modality may be matched to those in another. Thus a person may adjust the loudness of a sound in his ear to equal the apparent strength of 60-cycle-per-second vibration on his finger. When the vibration is changed, the observer changes the loudness to match the new apparent strength. The example in the log-log coordinates of Fig. 1 shows how the matching function turns out to be a straight-line power function whose slope (exponent) is given by the ratio between the exponent for loudness (0.6) and the exponent for vibration (0.95).

Reproduced from SCIENCE vol. 151, 4 February 1966, pp.530-541.

The basis for this result is as follows. If, given an appropriate choice of units, two modalities are governed by the equations

$$\psi_1 = \varphi_1^{\alpha}$$

and

$$\psi_2 = \varphi_2^{\beta}$$

and if the subjective values ψ_1 and ψ_2 are equated by cross-modality matches at various levels, then the resulting equal-sensation function will determine a relation between the two kinds of stimuli of the form

$$\varphi_1^{\alpha} = \varphi_2^{\beta}.$$

In terms of logarithms,

$$\log \varphi_1 = (\beta/\alpha)(\log \varphi_2).$$

In other words, in log-log coordinates the equal-sensation function becomes a straight line whose slope is given by the ratio of the two exponents. An interconnected net of exponent values has been validated by this direct matching procedure (5). On more than a dozen different continua the power-function relation has been confirmed and the value of the exponent has been checked by cross-modality comparisons.

This rather happy, if unexpected, development has demonstrated an important possibility. The stimulus-response relations for all the sense modalities can be mapped out without resort to numerical estimation on the part of the observers. The power functions obtainable by cross-modality matching make methods like magnitude estimation entirely dispensable.

On the other hand, magnitude estimation has the great advantage of convenience. Its prescription is simple. It calls for presentation of a series of stimuli in irregular order, if possible in a different order to each observer. The instructions may be modeled on the following example.

You will be presented with a series of stimuli in irregular order. Your task is to tell how intense they seem by assigning numbers to them. Call the first stimulus any number that seems to you appropriate. Then assign successive numbers in such a way that they reflect your subjective impression. For example, if a stimulus seems 20 times as intense, assign a number 20 times as large as the first. If it seems one-fifth as intense, assign a number one-fifth as large, and so forth. Use fractions, whole numbers, or decimals, but make each assignment proportional to the intensity as you perceive it.

Because not everyone is familiar with the concept of proportionality, it has sometimes proved helpful to start off with an experiment on apparent length of lines. The lines, six to ten in number, should cover a wide range of lengths—say, a ratio of about 50 to 1. After judging such lines in irregular order, most observers seem to achieve a reasonably firm grasp on the concept of assigning numbers proportional to magnitude.

The variability of magnitude estimations has been found to grow approximately in proportion to the magnitude, and to produce distributions that are roughly log normal. Consequently, averaging is done best by taking geometric means of the estimations. This method of averaging also has the advantage that, despite the different ranges of numbers used by different observers, no normalizing is needed prior to averaging.

Finnie and Luce say of the direct scaling procedure: "In addition to their theoretical interest in extending those methods and relations to areas other than psychophysics, knowledge of such a relation [a relation such as that of Fig. 3] can have considerable practical benefit. A magnitude scale on 10 or 100 items can be obtained from a group of people in, literally, a matter of minutes. The corresponding data for a Thurstonian analysis . . . requires more time to collect and is considerably more expensive to analyze, even with modern computation aids."

The ease with which a magnitude scale can be obtained from subjects instructed to match numbers to assorted items in a manner that preserves a proportional relation is indeed impressive, but a serious scaling venture will usually demand additional studies and may profitably be validated by other procedures, including cross-modality matching. As in any empirical inquiry, much depends on the level of accuracy required.

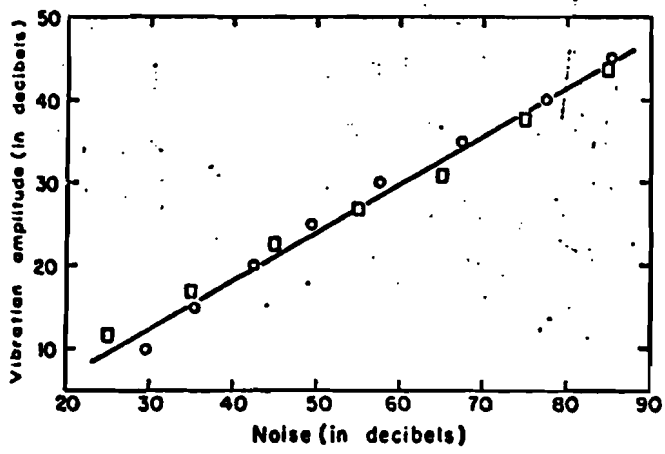


Fig. 1 (left). Equal-sensation function for cross-modality matching between loudness and vibration. The squares indicate that the observers adjusted the intensity of vibration on the fingertip to match the loudness of a noise delivered by earphones. The circles indicate that the observers adjusted the loudness to match the vibration. Each point is the decibel average of 20 matches, two by each of ten observers.

Esthetic Value of Handwriting, Drawings, Music

Two separate studies of the esthetic value of handwriting were carried out by Ekman and Künnapas (*17*). In both studies, samples of handwriting were scaled by Thurstone's method of pair comparisons and by a variant of the method of ratio estimation (*18*). The first study, in which seven samples of handwriting were used, is especially instructive because it failed to show a logarithmic relation between the scale by pair comparisons and the scale by ratio estimation. In the second study, 18 samples of handwriting were used, covering a wider range of quality. This second experiment demonstrated that the approximately linear relation of the first experiment became an obviously logarithmic relation when the wider stimulus range made the form of the function easier to determine.

A related study by similar methods was carried out on 17 drawings of a tree. The samples were selected from some 200 drawings produced by sixth-grade students (*19*). Again it was found that the confusion scale derived from pair comparisons was quite accurately proportional to the logarithm of the magnitude scale derived from ratio estimations.

Esthetic judgment in music was investigated by Koh (*20*), who presented 51 vocal selections and 60 piano pieces to various populations of subjects, including college students and patients in the alcoholic ward of a hospital in North Dakota. Each excerpt lasted about 15 seconds for the piano pieces and about 60 seconds for the vocal pieces. The total population of subjects, numbering 330, was divided into six groups, two groups each of college males, college females, and alcoholics. Half of the groups, one of each category, heard vocal selections, the other half heard piano selections. The subjects in each group made magnitude estimations of the affective value of each selection and also judged each selection on a category scale expressed in terms of nine adjectives ranging from "most pleasant," through "indifferent," to "most unpleasant." The category values were treated as a 9-point numerical scale, and the ratings for each piece of music were averaged.

For all six groups, the average rating scale value was approximately proportional to the logarithm of the geometric means of the magnitude estimations. The product-moment correlations between the category values and the logarithms of the magnitude values ranged from 0.90 to 0.96. For some groups there was a slight upward concavity of the curve; this is usually the case when category scales are plotted against the logarithm of the magnitude scale.

As Koh remarked, the relation between the category and the magnitude scales was strikingly invariant in spite of the differences in the age, sex, education, occupation, and pathology of the subjects. "These empirical invariances," he concluded, "strongly suggest the usefulness of magnitude estimation for complex judgmental processes."

Occupational Preference

Perloe measured the degree of prestige that attaches to each of 100 different occupations by means of two procedures: magnitude estimation and a 7-point category rating scale (*22*). The subjects were 40 undergraduates at Haverford College. Despite a certain mix-up about the instructions, the data show that the relation between the category scale and the magnitude scale is essentially the same for judgments of occupations as it was for judgments of musical selections in Koh's experiments. As with loudness, brightness, and other attributes for which there exists a stimulus metric, the mean category judgments define a scale that is almost, but not quite, a logarithmic function of the median magnitude estimations.

This is the expected outcome when the "noise" or variability in the experiment is large. When the variability is small, the category scale departs farther from the logarithmic form (*23*). Under more favorable circumstances—where, for example, the subject may be permitted to adjust a stimulus to bisect the apparent distance between two other stimuli—the partition scale may approach fairly close to the magnitude scale (*24*).

In the judgments of those of Perloe's observers who appear to have grasped the instructions, the range from the most to the least prestigious occupation was about 30-fold.

A roughly comparable range characterized the judgments of 74 students at the University of Stockholm who expressed their preference for 17 different occupations (*25*). In the Stockholm study two different procedures —ratio estimation and magnitude estimation—were used to scale occupational preference, and a third procedure —pair comparisons processed according to the assumption of case V—was used to produce a Thurstonian scale. The two magnitude scales were found to be linearly related, as had been expected. The confusion scale derived from pair comparisons approximated a logarithmic function of both magnitude scales.

The 17 occupations are shown in Fig. 6 in the positions assigned them by the geometric means of the magnitude estimations. Among university students in Sweden, the occupation of physician appears to be rated far out in front.

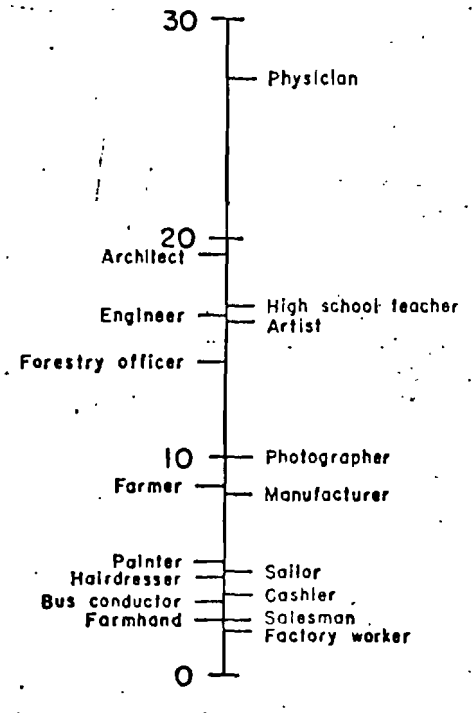


Fig. 6. Degree of preference for various occupations expressed by students at the University of Stockholm.

Comparable preferences for 30 U.B.C. students are found in the Appendix.

Seriousness of Offenses

On the basis of a preliminary study of opinion concerning a number of offenses, Ekman (*31*) selected descriptions of 17 more or less immoral actions for a study of moral judgment. The actions ranged from hit-and-run by a drunken driver down to stopping in a no-parking zone to mail a letter. Eighty subjects made pair comparisons and ratio estimations. As shown in Fig. 7, the pair-comparison scale based on a processing of the noise or confusion (case V) is very close to a logarithmic function of the scale based on direct ratio estimations.

Ekman's study, like most of those described thus far, was methodological in intent: the object of interest was method, not the achievement of a practical, substantive outcome. A full-length study in which method was the means rather than the end has been reported by Sellin and Wolfgang (*32*). Their 423-page book is directed at improvement of the methods used to compile police and court statistics for the purpose of measuring criminality in general and delinquency in particular. The research design placed major emphasis on delinquency events, not on delinquent persons, for the purpose was to measure the amount and type of harm to the community attributable to antisocial acts.

The general strategy of this 3-year study was as follows. First, a representative 10-percent sample of delinquency events was selected by random sampling from the universe of all such events in Philadelphia, Pennsylvania, recorded in the year 1960. Scaling procedures were then applied to events selected from the sample in order to convert the judged seriousness of the events into numerical scores. A final combination of all the information produced a delinquency index, a device that can be used to gauge the total incidence of delinquency and the effectiveness of whatever preventive measures may be brought to bear on the grave problem of antisocial behavior.

It is the second stage of the study that most concerns us here, the quantification of the gravity of delinquent acts, a quantification that must rest ultimately on the judgment of members of society. In brief outline, Sellin and Wolfgang proceeded as follows: A list of 141 offenses was first compiled, and a carefully phrased statement was made of each offense. These statements, typed on cards, were submitted for trial testing to 17 raters, mostly college students, who rated the seriousness of each offense on a 7-point category scale. Three representative offenses were then selected from each of the seven categories for use in further testing. These 21 offenses, presented in carefully randomized orders, were judged by 569 people—38 juvenile-court judges, 286 police officers, and 245 students from two universities. About half of each class of raters made magnitude estimations of the seriousness of the offenses; the other half rated the offenses on an 11-point category scale.

The next question of interest concerns the relation between the two kinds of judgments, category and magnitude. Figure 8 shows a direct comparison between the results for the 38 juvenile-court judges, 20 of whom used the category scale and 18 of whom made direct magnitude estimations. As is characteristic of prothetic continua, when the category scale of degree of delinquency is plotted against the magnitude scale, the curve is concave downward. When the same category ratings are plotted against the logarithm of the magnitude scale, the result is more nearly linear, but slightly concave upward, as shown in Fig. 9. Finally, when, instead of the averages of the category assignments, only the variability or confusion among the category assignments is used to generate a category confusion scale, the result is more nearly a linear function of the logarithm of the magnitude estimations, as shown in Fig. 10.

For different groups of raters, the age of the offender was specified as 13, 17, or 27 years, or it was left unspecified. Ten different plots like the plot of Fig. 10 were made from the judgments of the ten subgroups of raters. The impressive feature of the ten plots is their invariant form. When the total ordinate scale is taken as one unit, the slopes of the ten functions range from 0.22 to 0.31. The slope in Fig. 10 is 0.29. No significant differences were attributable to the age of the offender. It was the offense itself that seemed to determine the judgment of seriousness.

More important, perhaps, there was also impressive invariance across raters. Juvenile-court judges produced scales comparable to those produced by police officers and college students. It may be surprising that all three classes of raters agreed, for example, that stealing and abandoning a car is only about one-tenth as serious as robbing a man of $5 and wounding him in the process. According to the consensus, this latter crime becomes about two-and-a-half times as serious if the victim dies. Out of these magnitude estimations, say Sellin and Wolfgang (*32*, p. 268), "a pervasive social agreement about what is serious and what is not appears to emerge, and this agreement transcends simple qualitative concordance; it extends to the estimated degree of seriousness of these offenses."

The next major step in the study was an item analysis designed to refine further the statements used to define the offenses. The revised statements were used in a retest with a new population, a group of 195 students from still another university. This final testing gave results that correlated highly with the earlier data and thereby provided added justification for the construction of an index of delinquency based on a representative ratio scale of seriousness of offense.

An important feature of the final index is its provision for the additivity of offenses, a feature justified to a large extent by the outcome of the magnitude estimations of seriousness. Thus, the stealing of $5 is given a rounded value of 1. Breaking into a building also has the value 1. Breaking in *and* stealing $5 has the value 2, because the magnitude estimation score for the seriousness of the combined act was approximately double the estimate for each act separately. As another example, forcible rape has the value 11—8 for the forced sex act, 2 for the intimidation of the victim, and 1 for the inflicting of minor injury. The extraction of the additive components of the

complex delinquent acts was achieved, of course, through the process of analyzing the results of the magnitude estimations. It is doubtful that any of the raters would have been conscious of the underlying additivity in any explicit way, and some of them would probably be offended by the thought that one forcible rape can be equated to some number of money thefts. Nevertheless, both the quantitative estimates of large numbers of raters and the gradations in the punishments prescribed by law make a strong argument for equatability and additivity among offenses.

Punishment

How well does society's accumulated wisdom, or lack thereof, in legislating punishment accord with the judged gravity of offenses? In particular, what does the Pennsylvania Penal Code say about maximum penalties for the 21 offenses scaled in the main study? The answer is both interesting and encouraging. The product-moment correlation between seriousness of offense as judged by university students and maximum penalty stated in terms of time in jail was 0.88, provided a death sentence is interpreted as a jail term equal to the life expectancy of the median perpetrator of homicide. The correlation was even slightly higher, 0.94, for the magnitude-of-offense judgments by police officers. Both sets of results are shown in logarithmic coordinates in Fig. 12. As Sellin and Wolfgang express it (*32*, p. 327), "These correlations are surprisingly high considering the fact that the Penal Code provides no variation in the maximum [penalty] for amounts of money stolen and relatively few intervals between thirty days' imprisonment and death." Note also that the punishment scale is truncated at its lower end, for the smallest maximum penalty is 30 days in jail.

Another point of interest is the general form of the relation in Fig. 12. The straight line through the data represents a power function, because the coordinates are logarithmic. The slope of the line in these coordinates gives a measure of the value of the exponent. The slope (exponent) is clearly less than 1.0. Its value, 0.7, means that the penalty (time in jail) is not proportional to the seriousness of the offense. In linear coordinates the line in Fig. 12 would appear as a curve that is concave downward.

In order to have the full story on the justice of the maximum penalty specified by the Code, we would need to know another function: the judged seriousness or severity of various periods in jail—that is, the subjective value function for terms of imprisonment. Since that function was not directly scaled by Sellin and Wolfgang, we can approach the question indirectly by way of an assumption—the assumption that in this, the best of all possible worlds known thus far to *Homo sapiens*, the punishment fits the crime, provided both are assessed by direct subjective judgment. If that assumption holds, it follows that the subjective severity function for time in jail has an exponent that is the same as the exponent in Fig. 12—namely, 0.7. That exponent, with a value of less than 1, raises the interesting question whether people regard the severity of punishment that goes with various periods in jail as a decelerating function of calendar time. Is a sentence of 2 months less than twice as punishing as a sentence of 1 month? I would think so. If others agree, then perhaps the Pennsylvania Penal Code does indeed mete out roughly proportional justice.

In the preceding argument, the absolute value of the intercept of the hypothetical punishment function has been neglected. It will have to be considered, of course, before the complete story is told, because the *absolute* amount of punishment for a given offense merits as much concern as does the *relative* amount of punishment for different offenses.

However that may be, Sellin and Wolfgang have shown how to attack an urgent social problem with methods that were developed in psychophysics for the study of human sensory systems. The methods they borrowed have produced [illegible] useful results. It is a large and onerous task to develop and refine, by repeated revisions, a scale with useful properties in an area as complex as delinquent behavior. The one-shot experiment, so typical of the academic investigator, will not suffice when the goal is serious and substantial. The ratio-scaling methods used by Sellin and Wolfgang have a long history, and it is instructive that their development got its biggest push from a practical problem in acoustics (*35*). In the 3-year study of delinquency, the extension of the ratio-scaling methods to social variables has been dramatically achieved, largely because the challenge of the problem has justified the investment needed to track down and eliminate needless sources of noise and variability. Science seems to do its best when it faces a problem worth solving.

Functions for Individual Observers

Assessments of subjective value like those shown in Fig. 12 represent averaged results, and it may be objected that the scale of seriousness of delinquent acts does not represent the opinion of some particular person. Indeed it may not, for the first task must be to discover the consensus, if there is one. After a representative value function has been spelled out, it may or may not become profitable to ask about the exceptions. Whether, in a given domain, the consensus is sufficiently homogeneous to justify averaging is an empirical question, and one that has begun to receive attention in psychophysics.

It was for groups of observers that loudness and brightness were first found, on the average, to grow in proportion to the stimulus intensity raised to a power (*36*). As might be expected, little attention was paid at first to the question whether the power law would describe the reactions of each individual. Understandably, therefore, the question arises whether the ubiquitous power function may not be the result of group averaging. Pradhan and Hoffman (*37*), upon finding that some of their six observers did not produce power functions when judging apparent weight, concluded that individual functions "were found not to follow Stevens' law although averaging over observers does yield a power function. Stevens' power function thus seems to be an artifact of grouping."

That negative note contrasts with the view expressed by Ekman and Sjöberg (*38*): "After a hundred years of almost general acceptance . . . , Fechner's logarithmic law was replaced by the power law. The amount of experimental work performed in the 1950's on this problem . . . was enormous. . . . The power law was verified again and again, in literally hundreds of experiments. As an experimental fact, the power law is established beyond any reasonable doubt, possibly more firmly established than anything else in psychology."

An empirical answer to the problem of the individual function calls for the straightforward procedure of measuring and exhibiting the functions for individual observers. An early attempt to exhibit a collection of individual power functions was vetoed by a journal editor who pointed out, quite rightly, that nothing was shown by 39 straight lines in log-log coordinates that could not be summarized in a sentence or two (*39*). Since then, the problem of the individual function has led members of the Laboratory of Psychophysics at Harvard to develop new procedures. Lee McMahon made the first experiments on loudness with a technique that leaves the observer free to set the level of the sound intensity and also to estimate the loudness. It is a combined production-estimation technique which produces data that cannot be averaged, because each observer sets different stimulus levels and makes different estimates, at his pleasure. He may be asked simply to set as many levels as he likes and to assign numbers proportional to the loudness as he hears it. Results obtained with this technique by J. C. Stevens and M. Guirao (*40*) are shown in Fig. 13.

If the power law can be verified in a sufficient number of experiments, perhaps the point of interest will shift to the opposite side: What characterizes an observer whose judgments do *not* give a power function when the data are cross-checked by a variety of procedures? A single experiment or a single technique may not be enough to show conclusively that, for a given person, the perceived magnitude fails to grow as a power function of the stimulus magnitude. Almost anything may happen in a single experiment. A firm decision about an individual case may call for a multiple attack by estimation procedures, production procedures, and cross-modality comparisons. Hopefully, a multiple approach may even prove adequate to reveal an abnormal sensory function, such as auditory recruitment, in individual patients.

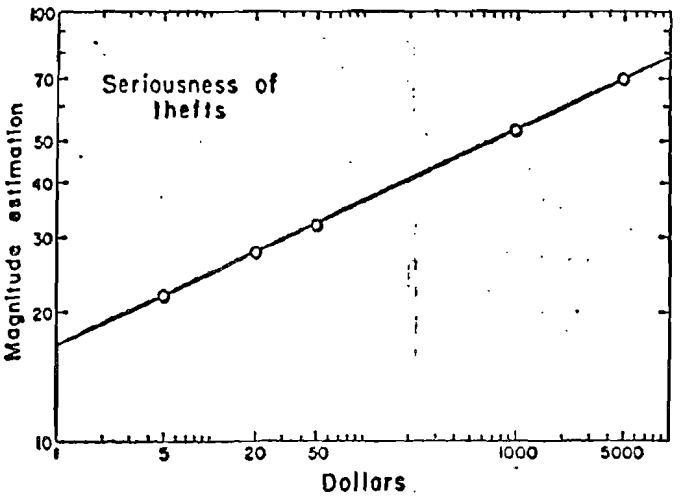


Fig. 11 (left). Magnitude estimations of the seriousness of stealing various amounts of money. The values on the ordinate are the geometric means of estimates made by 105 university students. The line, in log-log coordinates, defines a power function with an exponent of 0.17.

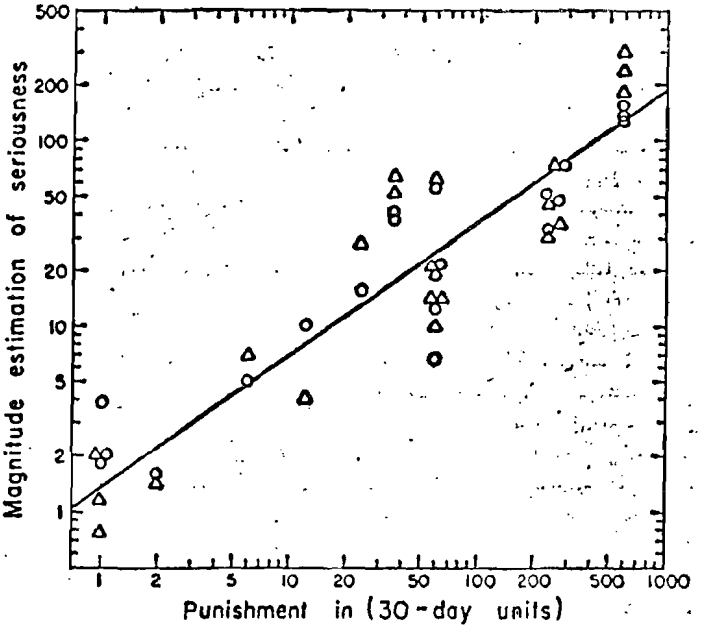
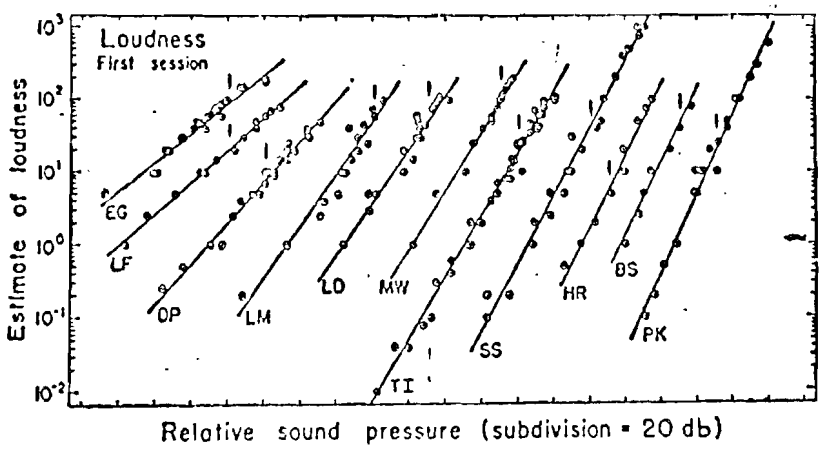


Fig. 12 (right). Relation between the geometric means of the judged seriousness of 21 offenses and the maximum penalty prescribed by the Pennsylvania Penal Code. The raters were police officers (circles) and university students (triangles). For plotting purposes the police ratings were multiplied by 0.5. The line through the data has a slope of 0.7.

Two features of Fig. 13 deserve comment. All 11 observers produced rather good approximations to power functions, but the slopes (exponents) varied from person to person. The analog of this second feature showed up also in the study of delinquency by Sellin and Wolfgang, where the range of the magnitude estimations for the seriousness of the offenses varied from person to person.

How should we regard these individual differences in range, or in exponent? Admittedly it would be something of a miracle if everybody's judgments followed exactly the same function for delinquency, loudness, or anything else. Perhaps the variations in how people use numbers and how they regard ratios are no more than the inevitable noise that characterize these complex processes. The fact that two of the lines in Fig. 13 have different slopes may mean that the two observers in question have different mechanisms at work in their auditory systems, but it may also mean that the two observers happen merely to differ about what they consider an apparent ratio. Further experiments may decide the point. In the meantime, there is growing evidence that the differences in the observed exponents, for a reasonable sample of observers, have one of the very important properties of noise—namely, the capacity to be averaged out. Note, for example, in Fig. 12 how nearly the average estimations by the police officers agree with the average estimations by university students. It is the stability of the function from group to group that makes the result useful.



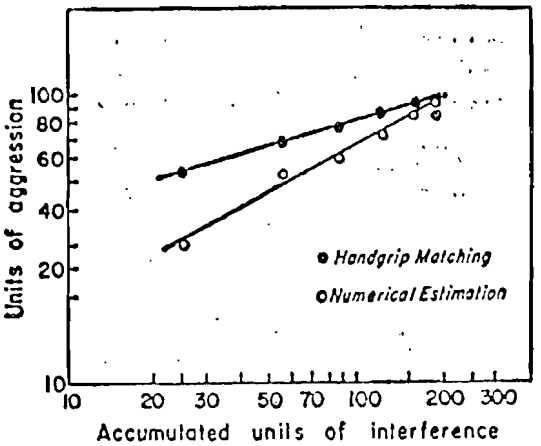


Fig. 13 (left). Individual loudness functions obtained from 11 observers (identified by initials) in the first session in which they both set the level of the stimulus and judged the loudness. The observers differed greatly in the range and number of stimuli they produced and estimated, but all the data approximate power functions. The small arrow above each function indicates a level of 80 decibels on the abscissa. Fig. 14 (right). Aggression, expressed as degree of dislike for a "leader" who interfered with the success of the subjects in a game, is plotted against the amount of the interference (number of points lost by reason of the leader's wrongheadedness). Forty subjects made magnitude estimations of their dislike and also squeezed a hand dynamometer by an amount proportional to the intensity of their feeling. The relative slopes of the two lines are predictable from other experiments.

Conclusions

What then are the invariances in these manifold experiments involving human judgment? A convergence of evidence from fields as disparate as psychophysics and criminology has pointed to stable and constant relations. One such relation states that subjective magnitude is a power function of stimulus magnitude. The underlying invariance then becomes the simple principle that equal stimulus ratios produce equal subjective ratios.

On many of the continua discussed above, the stimuli can be measured only on a nominal scale, for the stimuli are verbal statements, occupations, crimes, musical selections, and other nonmetric items. On those continua the power law cannot be confirmed directly, but there emerges another notable invariance.

For both kinds of continua, those based on metric stimuli and those based on nonmetric stimuli, there is a constant relation between the scale erected by direct judgment and the scale derived from a unitizing of variability or confusion. Whether the stimuli are measurable on ratio scales or only on nominal scales, the judgmental scale based on units of variability is approximately proportional to the logarithm of the scale constructed by one or another of the direct scaling methods. The extensive invariance of this logarithmic relation attests to a principle known throughout all of science—namely, that error or variability tends to be *relative*: the size of the error grows with magnitude. The principle finds expression under many phrasings: the standard deviation increases with the mean; the coefficient of variation remains constant; the signal-to-noise ratio stays put; accuracies are statable as one part in so many. The emergence of a similar canon in the subjective domain, a rule that variability tends to increase in proportion to the apparent magnitude, suggests an essential unity among the principles that govern quantitative relations in widely diverse endeavors.

For those who must build their science on one or another consensus of human judgment, a way seems open for an effective quantification.

Stevens' list of references and notes appears in the appendix.

Second Opinion

Norman and Lindsay's text on Human Information Processing summarizes many of the essential psychophysical findings obtained through Magnitude Estimation techniques and comment on the effectiveness and the convenience of the technique thus:

> After years of experience with magnitude estimation as a tool for measuring subjective experience, it would appear to be a reliable, robust method. It is simple and effective. It gives reliable answers, so reliable that it can be used in a class as a demonstration of scaling without any fear that the answers will come out wrong. In fact, the main difficulties with the method come when the experimenter tries to help it along: He suggests that the subject might limit the numbers he uses; or he presents the standard over and over again, lest the subject forget what it was like; or he collects many trials of responses to get a good statistical reliability. All of these improvements make things worse. As Stevens says, "He should keep hands off and let the observers make their own judgements." In fact, it is not even necessary to present the standard. Let the subject assign whatever numbers he feels appropriate to each stimulus as it is presented. This is much more natural for the subject than arbitrarily telling him what number he should use for that first one. Actually, today most experimenters do not use numbers. They use instead a procedure called *cross-modality matching* (described a couple of sections hence).

In many ways, it is precisely this laissez-faire attribute that makes Magnitude Estimation a particularly attrative means of collecting subjective value data in the Environmental/Energy domain. For information so "unconstrained" as subjective values of environmental impact, it would seem that (as perceived by the survey respondents) the less formalized the data collection strategies, the more likely that technique and value will remain unconfounded. A brief "Gedankenexperiment" will corroborate this: If you imagine a scale (vector) along which the "Riskiness" of a half dozen hazardous events are ranged, it is far easier to visualize the spacings between those events all simultaneously on that simple scale than if each pairwise combination is hedged against probabilistic odds to achieve an indifference curve. The importance of the simultaneity feature will become evident in the treatment of cross-modality matching.

The power functions for 14 stimulus/judgement relationships are shown in the following table. Notice how the power functions increase roughly for stimuli as the associated physical energy form becomes progressively more dangerous to the organism. Loudness, brightness and odours would need gigantic quantities of stimulus energy present to be hazardous; whereas shock, physical force, heat and heaviness probably span only a few orders of magnitude from threshold to death.

Table 3

Judgment	*Power*
Loudness (one ear)	.3
Brightness, dark-adapted eye, target of 5°	.3
Smell of coffee odor	.55
Taste of saccharine	.8
Taste of salt	1.3
Taste of sucrose	1.3
Cold (on arm)	1.0
Warmth (on arm)	1.6
Thickness of wood blocks as felt by fingers	1.3
Heaviness of lifted weights	1.5
Force of a handgrip	1.7
Loudness of one's own voice	1.1
Electric shock applied to fingers (60 Hz)	3.5
Length of a line	1.0

It is tempting to speculate how clearly that same relationship will emerge when data are available which link subjective value perceptions with demonstrable physical magnitudes. For instance, what power function exists between:

1) Judgements of perceived risk and number of lives potentially affected
2) Objections to a reprocessing plant and land area needed
3) Objections to a reprocessing plant and extent of steam plume produced
4) Objections to a reporcessing plant and number of marine species affected
5) Objections to a reprocessing plant and remoteness from the survey respondent

The critical question is, of course, whether one can presume that because such power functions obtain between specific physical stimuli and corresponding judgements magnitude, that they also might obtain between subjective stimuli and judged magnitude.

The mathematical tractability (to use Baecher's term) of Magnitude Estimation is formal, simple and well developed. Further (and simpler) to Steven's own treatment earlier, Norman and Lindsay illustrate with three physical stimuli:

Interpretation of the power function

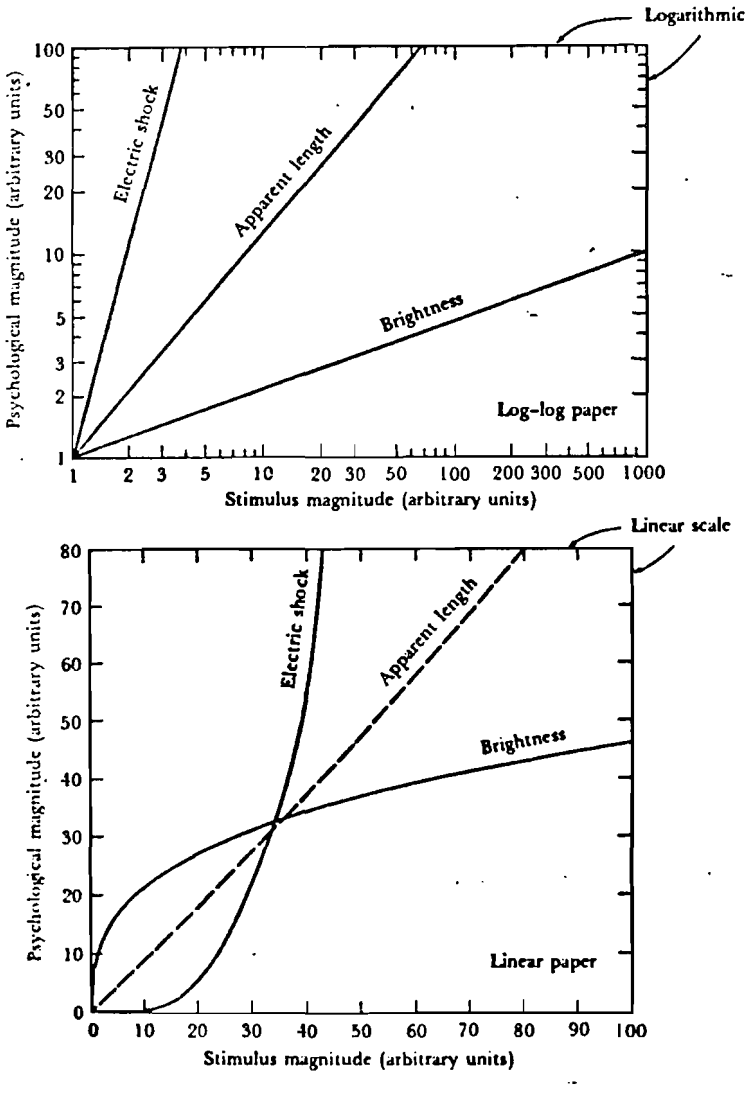
The power function is simply stated as $J = kI^p$. By taking logarithms of both sides of this equation, we find that

$$\log J = p \log I + \log k.$$

This is a simple result. It means that plotting the logarithm of psychological intensity on the vertical axis of a graph and the logarithm of physical intensity on the horizontal axis gives a straight line which has a slope of p and an intercept of $\log k$. Alternatively, the points can be plotted on the special graph paper which has both axes stretched out logarithmically—the graph paper which is called log–log paper. This simple relationship makes it easy to test the power function: When the results are plotted on log–log paper they ought to lie in a straight line.

FIGURE A-8

From Stevens (1961a).



The only restriction seems to be that the psychological dimension (criterion) be additive -- even if that additivity consists of many events summating.

Range of applicability

It is possible to use magnitude estimation procedures to judge almost any psychological dimension that is additive (or prothetic). Sellin and Wolfgang (1964) used this technique as a tool for measuring the way that society viewed the seriousness of crimes and of punishments. For example, subjects (juvenile court judges, police officers, college students) rated the seriousness of crimes. They judged that stealing and abandoning a car is .1 times the seriousness as robbing a man of $5 and wounding him. The robbery increases in the seriousness of the crime by a factor of 2.5 if the robbery victim is killed. Ratings on the seriousness of a robbery as a function of the amount of money stolen produced a power function with an exponent of .17. Thus, in order for one crime to be considered twice as serious as another, about 70 times the amount of money must be stolen [$70^{.17} = 2$].

Figures A-4 & A-5

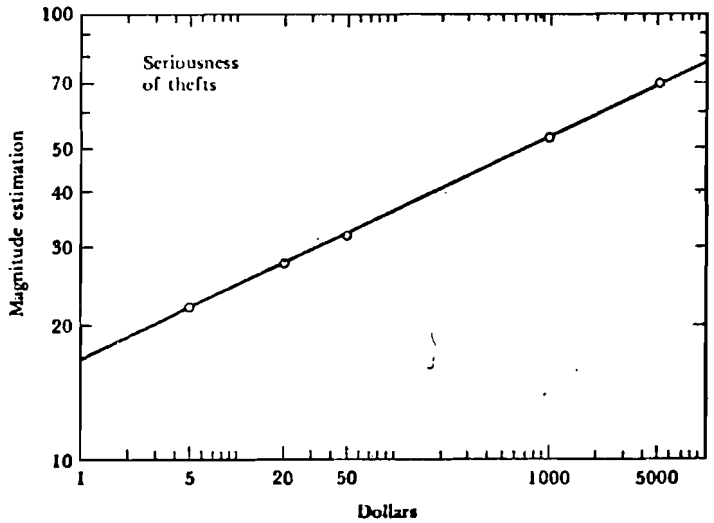


FIGURE A-9
From Stevens (1966a).

The activity feature illustrated in Steven's assessment of the components which contribute to perceived seriousness of rape, coupled with the robbery example above illustrates how powerful a technique Magnitude Estimation would be in quantifying the public perception of the risk of nuclear power generation and fuel reprocessing, since the occurance of a serious accident almost surely entails the summation of several quite specific components. In particular, Magnitude Estimation affords the chance to see just how large the "nuclear spectre" component is in assessment of any malfunction.

How many psychological units of seriousness must be assigned to any accident simply because it happens to occur in a nuclear station even though no nuclear substances or equipment may be involved?

The hope is that repeated monitoring at two or three year intervals will show the "nuclear spectre" component approaching zero as public education and familiarity divest nuclear energy production of its "bogey" image. Surely in the United States, at least, AEC imposed secrecy and public exclusion have long served as a powerful cue that "something dangerous and risky must be going on".

Cross-Modality Matching

One difficulty with magnitude estimation is its reliance on numbers. How does one assign a number to a sensation? Look at the brightness of this book page. What number do you assign to it? 10? 1000? 45.239?

A simple way to avoid the criticism, however, is simply to avoid the use of numbers. The easy way is to have someone judge the subjective magnitude of one event by producing an outcome that he feels is equal in subjective value.

One simple method is to have a subject listen to different sound intensities, say, and tell us their loudness by squeezing his hand as hard as he feels the sound to be loud. We measure the squeeze pressure with a dynamometer. Alternatively, we could have someone adjust the intensity of a tone until it sounded as loud as a light was bright, or draw a line as long as sandpaper was rough, or adjust an electric shock to have the same psychological magnitude as the strength of the odor of coffee. Does this seem a strange method to you? Try it (see the experiment described later in this section). The description is strange, but in practice it is quite simple and direct.

We can predict what the results of these cross-modality matches should be. Let us compare two continua, A and B. We do standard magnitude-estimation experiments for each, finding that for intensity values of I_A and I_B, the judgments of psychological magnitudes J_A and J_B are represented this way:

For estimates of A:

$$J_A = k_A I_A^a.$$

For estimates of B:

$$J_B = k_B I_B^b.$$

Now, if we ask our subject to observe a signal from A which has intensity I_A and produce a value of intensity on B, I_B, so that the two psychological impressions are equal, we know that

$$J_A = J_B,$$

and so

$$k_A I_A^a = k_B I_B^b.$$

Thus, if we solve for the value of I_B necessary for the judgment of B to match that of A

$$I_B^b = \frac{k_A I_A^a}{k_B}$$

and, taking the bth root of both sides,

$$I_B = k I_A^{a/b}, \qquad \text{where} \quad k = \frac{k_A}{k_B}.$$

Thus, we still get a power function when we plot the intensity of B that the subject claims matches the subjective impression of the intensity of A. The exponent of the power function obtained by cross-modality matching is given by the ratios of the exponents which we get in a magnitude estimation experiment.

It is difficult to overestimate the importance of this matching feature of Magnitude Estimation. Although it seems obvious that psychological continua might be represented in terms of each other, it is quite something else to demonstrate that indeed they are. The term most often used for this cross-modal representation is "synesthesia". The creative artist who attempts to join two media, music and dance, music and theatre, poetry and music synthesizes the reinforcing attributes of each of two or more modalities in his work to achieve the fusion of artistry that is choreography, opera or song. Doubtless, cross-modality matching is the atomistic analysis of all art form.

The thing that makes such matching so attractive is that trade-offs among psychological criteria can be made without the necessity of intervening numerologies, be they Magnitude Estimation, odds, indifference curves or probabilities.

Just how clearly the values predicted by cross-modal matches compare with the measured values can be examined for several physical continua in this table. Differences between predicted and obtained exponents do not exceed .07 and average about .03.

Table A-4 The Exponents (Slopes) of Equal-Sensation Functions, as Predicted from Ratio Scales of Subjective Magnitude, and as Obtained by Matching with Force of Handgrip[a]

	Ratio Scale	*Scaling by Means of Handgrip*		
Continuum	*Exponent of Power Function*	*Stimulus Range*	*Predicted Exponent*	*Obtained Exponent*
Electric shock (60-cycle current)	3.5	0.29–0.72 milliampere	2.06	2.13
Temperature (warm)	1.6	2.0–14.5°C above neutral temperature	.94	.96
Heaviness of lifted weights	1.45	28–480 gm	.85	.79
Pressure on palm	1.1	0.5–5.0 lb	.65	.67
Temperature (cold)	1.0	3.3–30.6°C below neutral temperature	.59	.60
60-Hz vibration	.95	17–47 dB re approximate threshold	.56	.56
Loudness of white noise[b]	.6	59–95 dB re .0002 dyne/cm²	.35	.41
Loudness of 1000-Hz tone[b]	.6	47–87 dB re .0002 dyne/cm²	.35	.35
Brightness of white light	.33	59–96 dB re 10^{-10} lambert	.20	.21

[a] From Stevens (1961a).

[b] There is a technical issue here that often causes confusion. We specified that the exponent for loudness judgments as a function of sound intensity had a value of .3. Yet the table shown here lists the exponent as .6. Why the discrepancy? The answer is simply that sound is measured both in units of energy and amplitude. Sound *intensity* refers to energy measurements; *sound pressure level* (*SPL*) refers to amplitude measurements. Sound energy is proportional to sound amplitude squared ($I \approx A^2$). Hence, if we write the power function, we find that

$$J \approx I^{.3} \approx (A^2)^{.3} \approx A^{.6}.$$

Both exponents are correct: .6 applies when sound pressures are measured; .3 when sound intensities are used.

Similar graphic representations for several more psychological dimensions show that each of these dimensions can be accurately matched to Force of Handgrip or setting Sound Pressure levels, without the intermediate necessity of representing the stimulus as a number or a probability.

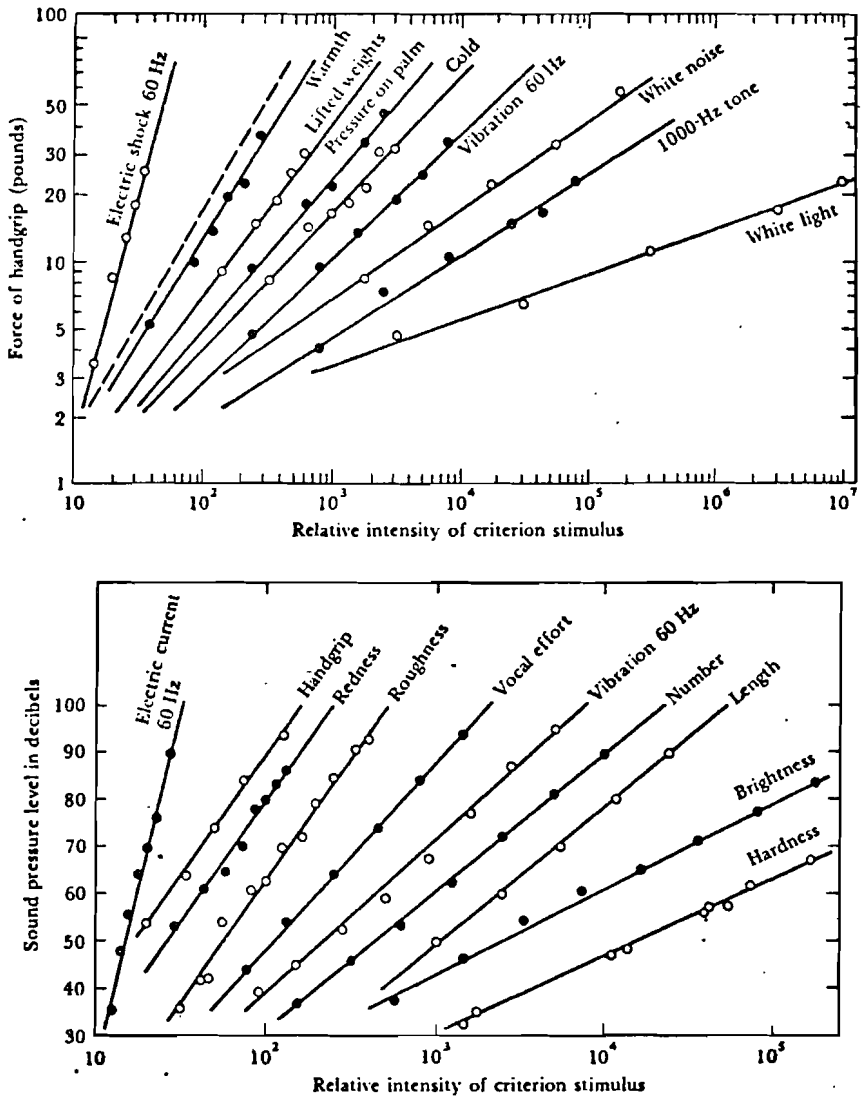


FIGURE A-9 *Top: Equal-sensation functions obtained by matching force of handgrip to various criterion stimuli. The relative position of a function along the horizontal axis is arbitrary. The dashed line shows a slope of 1.0 in these coordinates. Bottom: Equal-sensation functions obtained by matches between loudness and various criterion stimuli. The relative positions of the functions are arbitrary, but the slopes are those determined by the data. From Stevens (1966d).*

Why precisely is this cross-modal matching feature so noteworthy? The reasons come from several directions; some of them inherent in the nature of the phenomena to be assessed, some in the assessment strategies we use, others in the inferences we wish to make from the numbers that are the tools of analysis.

Range of Convenience

One critically important feature of the Values I.I.A.S.A. wishes to explore is the range of magnitude to be included. Consider the span of things a nuclear power plant should do, might do, and certainly should not do. It should generate steady, controllable power (p = .99999), it might require occasional maintenance (p = .10), it might even break down occasionally (p = .0001), but it certainly should not explode (p = 10^{-6}- 10^{-8}). Now in that small exercise alone, a range of 7 or 8 orders of magnitude has been specified. Consider (on an orthogonal axis) the numbers of people directly involved by those 4 occurences. Operating properly, some 10^{6}people may be served, occasional maintenance may require as many as 10^{1} people, a breakdown could involve 10^{2} emergency and standby personnel (including those who would balance the power load from other sources in the net) and an explosion could harm or kill 10^{2}-10^{4} people. Thus, another 6 orders of magnitude.

Now one might describe each of these events as "Exceptionally Desirable" or "Profoundly Undesirable" at the extremes, and "Routinely Desirable" that the plant perform with a minimum of maintenance; but it is highly unreasonable to expect words to be an exact reflection of the judgemental assessments of events whose range spans 6 to 8

orders of magnitude. Even with the number of verbal modifiers which one can muster to graduate between "profoundly pro" - and "profoundly con -", it is doubtful that one could reasonably expect more than 10 or so. Thus we should be requiring some 10 verbal distinctions of the judgemental dimension to adequately map onto 6 to 8 orders of magnitude! Even Lotfi Zadeh might object!

FRACTILES

Fractile and Indifference Approaches

The immediate response to the inadequacies of verbal distinction is a probabilistic one. Since probabilities range smoothly between zero and unity, we merely assign an appropriate p - value and the problem disappears. Or no? In the vigorous sciences where the probability of an event is known (or calculable) this may well work better than in this present case where some fix on hypothetical events is also desired. Such events have no known p - values.

Indifference

The usual strategy is to approach the dilemma by one of two strategies, either the fractile as the indifference routes. In the indifference case, the respondent is presented 2 events and must assign probabilities to each (sometimes one is prespecified) such that his assessment of event A multiplied by its likelihood (P_A) makes its payoff equal to the assessed value of B x P_B. In essence, the payoff matrix of the two events makes the respondent indifferent to which game he plays.

That is surely a step forward. A tenfold increase in the discriminating power of the judgemental dimension has occured. (100 probability points vs 10 words). But in another sense, the most interesting portions of the assessment curves have been lost. What the investigator

is left with is the equivalency between two events of their zones of disinterest. In actuality, the procedure loosely fixes the mid-points of the two curves, in that the midpoint must lie somewhere in the area of the product of the two indifference widths.

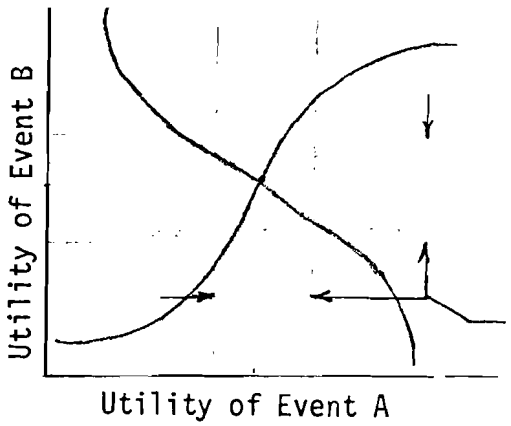


Figure 10

Fractiles

Fractiles, of which indifference strategies are a special case, can map one curve onto another at as many fractile points as one might care to choose. At this point, the full double digit resolution of the perceived value of event A can be experienced in terms of Event B.

Fractile strategies in one form or another can probably deal with most assessment needs up to the limits of respondents abilities to express judgements as numbers on probability weights.

Numbers

Just how much different people's assessments of a single event collected via Magnitude Estimation vs cross-modal matching routes is not exactly clear. Neither Stevens nor his co-workers (to my knowledge) have approached the question directly, although Stevens does comment on the slight differences which occur when people assign stimuli to pre-arranged numbers, rather than numbers to stimuli.

Fractiles allow double digit judgemental resolution to be mapped onto stimuli which may span 6 - 8 or perhaps even 10 orders of magnitude. While that is rather coarse mapping, there is a real question whether anyone really understands the range over so great a span to use even powers of 10 intelligently. Surely the means are available to divide the judgemental dimension into as many graduations as one might wish, but whether respondents would ever use more than a few of them, coarsely chunked, is conjectural.

Images

An attendant problem was raised by Haefele during the summer: How would an experimenter convey to a respondent stimuli or events in sufficiently vivid fashion to enable intelligent scaling of requisite magnitude span and clarity of resolution?

For physical stimuli, the problem vanishes, one presents weights, sheets of graded sandpaper, line segments, or sounds tones, flashes lights or administers shock.

Most precedents for less physicalistic stimuli have already been described in the extensive quotes. One types clear, concise descriptions of events on cards and directs the subjects to read the description and assign appropriate numerical magnitude, one after another without referring back to earlier cards. My own students viewed one or two word names of professions, and yielded the curve shown in the appendix.

Conversely, Indow presented pictures and descriptions of watches; Ekman and Künnapas presented handwriting samples; Koh played recordings of vocal and piano selections. Elsewhere odours, masculinity and feminity, agression and political conservatism have been investigated.

A common sense method for generating or classifying descriptions of events is presented in the next section. Whether the events should be photographed, sketched, or otherwise simulated is open to inventiveness. If Stevens is correct however, the ratios among events, whatever the presentation medium, will be invariant.

ASSESSING EVENTS AND CRITERIA

Criteria: Implicit or Explicit

Whatever the particular complement of measurement and decision techniques, the two guiding questions are: How are various events assessed, relative to each other?, and Along what criteria are those assessments made? In principle, this is a simple enough problem: How much? and In terms of what? Yet the examination of these problems have concerned psychometricians since the mid 1920's and the issues are not yet fully resolved.

Suffice it to say that there are two schools of thought; both have merit and both have difficulties. The one school holds that persons are really incapable of assessing multiple events in any more precise sense than to say "A is more similar to B than it is to C." A matrix of event comparisons would simply comprise a set of comparative Psychological Distances. This same school holds that if there are any such things as criteria, they will be discoverable as clusterings in the distance matrix, but only after the fact.

The other school replies. "Nonsense! I am quite capable of discerning that two events are very similar in terms of cost, but quite different in their esthetic appeal." And so the argument goes.

Such meetings-of-the-mind as even occur between the schools usually happen in or near the camp of the hypothetico-deductivists who argue to the first school "What you claim may be true, but you

you must surely admit that the choice of events to include in the first instance was biased by the ways you felt the events might differ or coincide." To the second school (criterion-explicit) the H-D's argue "you may surely assign an event to some scale positions on 2 or more criteria, but whether the event more clearly aligns with Criterion 1 or Criterion 2 is a matter for experienced verification." Something approximating truth emerges from the iterations between the two.

A study not too long past unsettled everyone's pet dogmas by testing 5 strategies for assigning events to criteria, including random assignment. Disturbingly (for some, at least) all four non-random schemes appeared to be equally useful and to account for about equal amounts of criterion variance, even though overlap among the 4 was next to nil. This suggests that any reasonably sensible framework for investigating psychological scaling has at least even chances of making sense out of the data. (Jackson, 1971)

What then are the event catagories that IIASA should consider? Here is a starter list, but a very first task, should be to convene a brainstorming session and make the list exhaustive. In this session, all non-trivial items should be submitted, critical assessment and classification attempts are strictly out of place until the category pool is formed. IIASA should examine events in these categories:

Risk
Cost
Benefits
Environmental Degradation
Landscape Esthetics
Waste Heat Problems
Waste Products
Cheaper Energy
Increased Energy Demand
Energy Conservation
Transition Fuels
Land Use Issues
Energy Distribution
Energy Storage
Central vs Local Energy Production
Infrastructure Momentum
More Efficient Use
Labour and Manpower Needs
etc.
etc.

Once a hundred or more such categories have been generated, the second step is to sort them, however arbitrarily, into logical groupings. At this stage some weeding out of redundancies, trivial categories, etc. is appropriate, but no discards (yet). You will save your collective selves considerable time and duplicate work if you will record the first list one each per individual slip of paper, so they can be shuffled around later.

Lists of this sort on looseleaf sheets of paper in people's personal notebooks somehow never achieve the full attention focus of all group members. Keep them public and keep them manipulable.

Progressing from Categories to Events:

This stage is the most critical one and frankly, the most difficult too. A category is not an event! And yet it is events that are to be assessed, by whatever means. The entire survey strategy can be ever-so-well planned, but if the specific events people are to examine are ill-defined, fuzzy or even non existant, only chaos can result.

My experience suggests that the best route at this point is common hard work. Each event category should be examplified by him to 7 specific instances. Whether these instances are gleaned from accident files (as for risks) or whether they are specific, concise, plausible and concrete. The simplest strategy for each category is to charge each team member to manufacture 3 such instances for each category. But the members must work individually and independently of each other until the events are all generated. Believe

me, group efforts do not succeed here.

If each event category has been recorded on a 3 x 5 card, then all such categories sorted into logical groups and filed (so arranged) in a card file, thus the event generation task is greatly facilitated. Each team member takes the card file (which should contain 500-1,000 cards by groups) into a seculded room and proceeds from front to back of the file writing 3 event descriptions for each card in the file. (Be smart and use different card colours for the categories then for the events). The resultant event collection could easily be 6,000 to 10,000 cards if there are 3 to 5 team members.

Clearly, a well-done media search can supply a very sizeable fraction of these events as news reports of real occurrences.

Sharpening the Events

If the event-generation task is done carefully and seriously, there should be 10-12 examples for each event class. From this number, by selection and combination, 5 clear, concrete and vivid examples can be produced. The number of final events may number about 2,000-3,000.

In sequence, the strategy is as follows:

Brainstorming Event Categories
Grouping the Categories
Refining Groups (weeding, elaborating, sharpening group definition)
Generating 3 Specific Events for each Category (in each group) per team member
Refining the Events List

When this is all complete, a file of a few thousand events will be the result - - ordered according to this format.

Group Characteristic A

Category Name 1
Event 1
Event 2
Event 3
Event 4
Event 5
Category Name 2
Event 1
Event 2
Event 3
Event 4

Group Characteristic B

Category Name 1
Event 1
Event 2
Event 3
Event 4
Event 5

It should be no surprise that this method produces a list of events whose Group Characteristic is the approximate criterion on which the events can be scaled. Thus, whether one wished to proceed according to either assumption about criterion - - implicit or explicit - - the events to examine are now ready made.

Discounting in Space/Time

One specific hypothesis of this study will examine the rate at which people discount the percieved impact of an event as that event is further and further removed from them in space and in time. This necessitates the need of one further (double) permutation of the event list - - that each event (or at least a controlled sample) be allowed to accumulate systematically manipulated times and

locations, perhaps thus.

Part Time: Event X Occurred
- Yesterday
- Last Week
- Last Month
- A Year Ago
- 5 Years Ago
- 10 Years Ago
- 20-25 Years Ago
- 50 Years Ago
- 100 Years Ago

Future Time: Event Y Will Occur
- Tomorrow
- Next Week
- Next Month
- etc.

Space: Event Z Occurrs
- In Town (10 km)
- Next Town (100 km)
- In City A (1,000 km)
- On Continent B (10,000 km)

Clearly, one would phrase events so as to be non-obvious and unobtrusive about the permutations.

CRITERION	RISK
Event Class I	Explosions
Event	An Entire Reprocessing Plant Explodes
permuted in time	. . . today, next year, 5 years, 10 years 20 years
permuted in space	. . . in town, 10 km. away, 100 km. away, 100 km. away
permuted in impact	. . . and 10, 100, 1,000, 10,000 people are killed
Event	A Steam Turbine Explodes in a Reprocessing Plant
permuted in time	. . . 1 year, 5 years, 10 years, 20 years, etc.
permuted in space	. . . in town, +10 km., +100 km., +1,000 km., etc.
permuted in impact	. . . killing 1, 5, 10, 100, 1,000 people, etc.
Event	A Steam Turbine Explodes in a Coal Powered Plant
permuted in time	. . .
permuted in space	. . .
permuted in impact	. . .
Event Class II	Leaks
Event	A Storage Tank Leaks
permuted in time	. . .
permuted in space	. . .
permuted in impact	. . .
Event	A Transport Vehicle Spills Enroute to the Plant
permuted in time	. . .
permuted in space	. . .
permuted in impact	. . .
Event	Sewage Discharge from the Plant Becomes Contaminated
permuted in time	. . .
permuted in space	. . .
permuted in impact	. . .
Event Class III	Materials Unaccounted For
Event	Ten Micrograms of Material Cannot Be Accounted For
permuted in time	. . .
permuted in space	. . .
permuted in impact	. . .
etc.	

The Event Generating and Classifying Matrix:

These preceeding pages can be shown in model form with the single Category Name (Criterion) of risk shown as an example. In this instance, 9 time permutations are shown, together with 4 spatial and 4 impact permutations. Consider this matrix illustration only, but it does convey the ease with which a serious of events can be made to span an immense range of convenience. Additional criteria would extend to the right, while subsequent event clauses would read downward.

To make the actual number of events small enough to be attempted by any one person in any one session, some fairly mundane event sampling must first occur. Otherwise the total number of events generated by a card file of only 1,000 events would total:

	1,000
x	9 time permutation
x	4 space permutation
x	4 impact permutation
	96,000 items to be assessed!

Sensible respondents would refuse and even willing ones couldn't manage.

Stratified sampling of the event list would ensure that any one respondent would receive events ranged in space on time or impact, but not radomized across, since to measure decay of perceived impact our time means that several respondents must examine the same event at several time intervals.

Some Exploratory Issues

A host of topics suggest themselves once a rich pool of events is available as the exploratory tool. But numerous as these topics are, even they can be sorted into some logical groupings. Broad categories of exploratory directions would include:

Differences among Special Interest Groups

Trade-offs among criteria

Placement of Specific Events along one or more criteria

Impact Discounting over Time

Impace Discounting over Space

Styles of Non-rational or Counter-intuitive Assessment

The Generating Matrix (with its permutations) allows exploration of all these issues in relatively rigorous fashion, given that some reasonable number of people who represent public views have participated in the assessment survey.

The following are some notes and thoughts on several of these topics, although in a quite different order.

Tradeoffs among Criteria:

To the Energy Project, this is likely the most compelling issue. Specific questions in this domain might be:

How many dollars is a hectare of landscape worth?

How many dollars is one degree cooler waste water worth?

How many BTU's is today's dollar worth? Tomorrow's? 10^5 km away's?

How many lives is today's energy production worth?

How many lives is each Risk Level worth?

How many BTU's of energy produced (or not produced) is a hectare of Landscape Worth?

How many Risk units is public recreational access to the energy installations worth?

Clearly this list could go on and on (and probably should, in quite an explicit sense). In short, what is really needed in a graph like that on page 43, where the abscissa is measured on dollars on BTUs and the curves are all the other criteria, both hard and subjective which command concern. See Figure 10 for a fictitious example.

Again, one significant advantage of the Magnitude Procedure is that the assumption at the outset is that matches between two criteria are likely not linear, therefore rule-of-thumb figures need not be assumed constant over the entire range of two matches. For example, in cold blooded terms, a human life is worth about what insurance companies will reimburse for one, around $30,000. Corporations are insured against damages for about this amount, court settlements in cases of accidental death approximate this figure too. Does this mean that if a nuclear accident should kill 500 people, that the social value of those lives is 500 x $30,000? Probably not, since in this, as in most other comparable matches, there is a function of decreasing marginal utility. Stevens shows that thefts of various amounts of money exhibit a decreasing function of about $\$^{.17}$. Each of the criterion matches one might select to investigate would exhibit a power function of its own.

In summary, the ultimate goal would be to produce an equivalency matrix in which the cell entries were the power function linking any possible pair of criteria, hard, soft or hypothetical, as suggest in Figure 11 . Any pair in

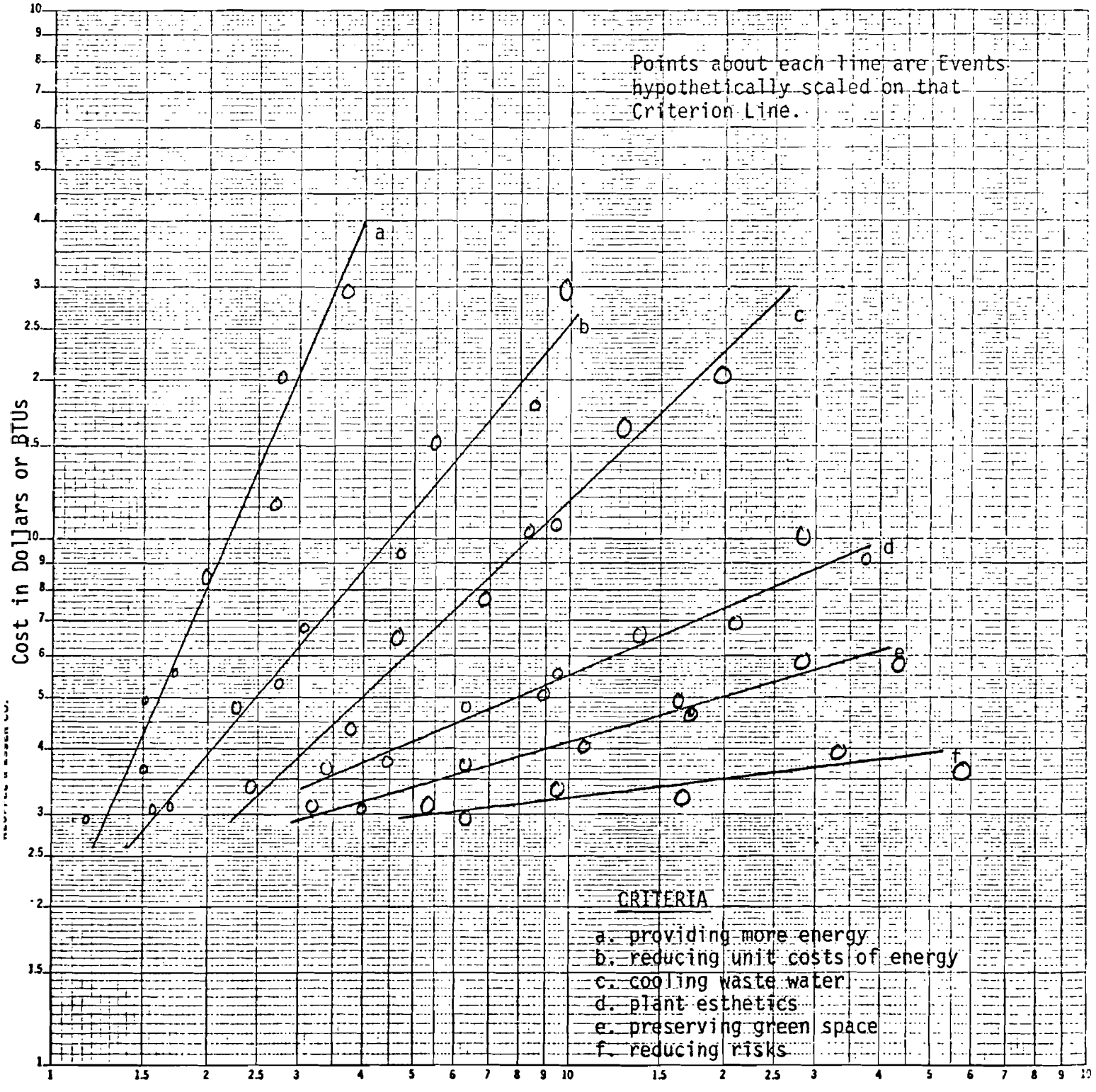
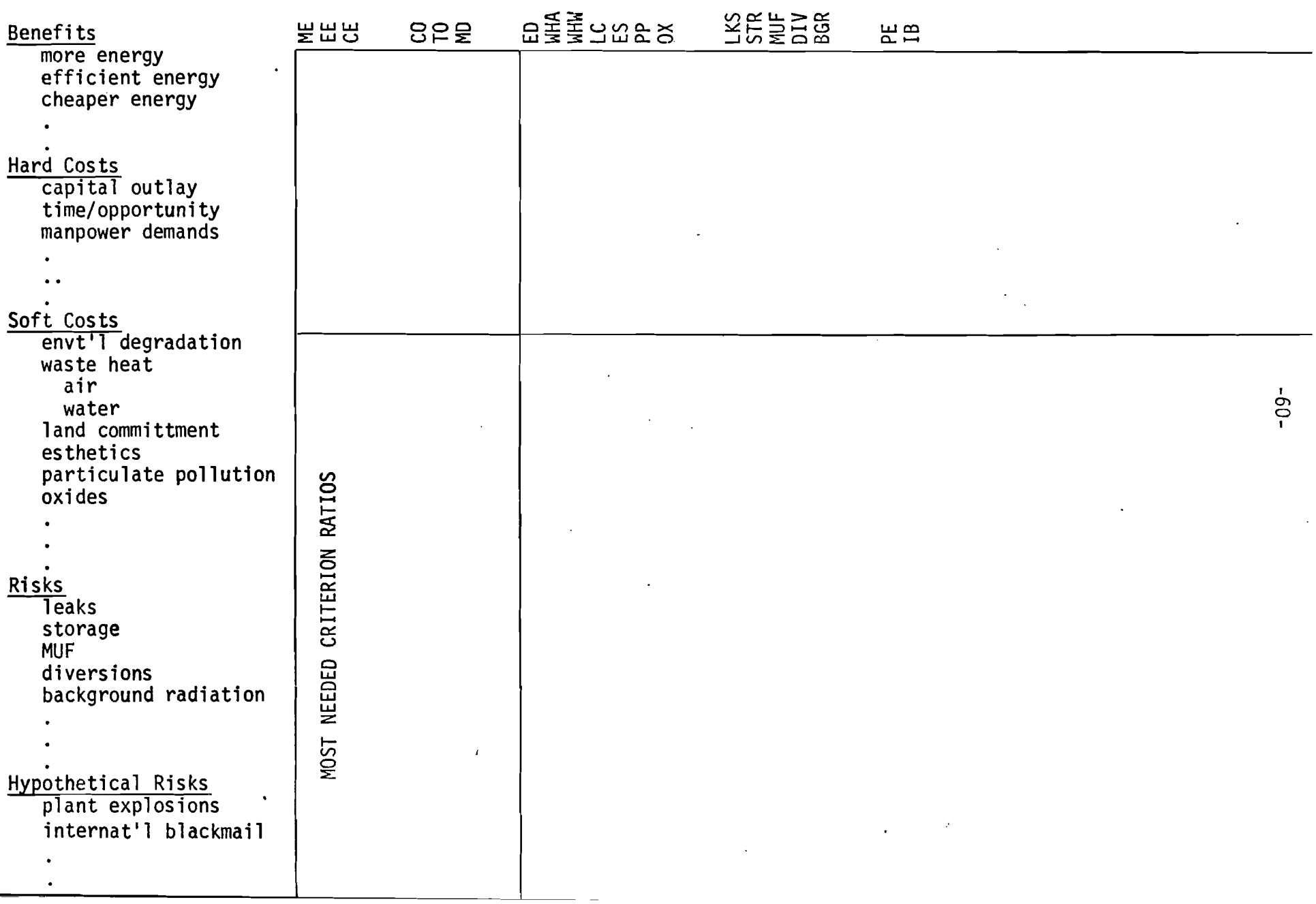


Figure 10.
Hypothetical curves and units for costs of several subjective values

Figure

A MATRIX OF CRITERION TRADE-OFFS



the matrix would be interesting, but the entries in the subsection labelled Most Needed Criterion Ratios are just that.

Discounting in Time and Space

The term "discounting" has been used throughout in a loose sense to refer to the decreasing marginal impact of an event (real or hypothetical) as it is perceived at greater and greater distance in time or in space. The concept is directly analogous to the opportunity cost of money: a dollar invested now is worth more than the same dollar invested next year or five years hence, thus; an event occurring now, has greater impact than the same event occurring one, five or ten years hence. Probably it also has greater impact than the same event one, five or ten years ago, too. If so, the monetary analogy suffices, however. Similarly an event occurring right here, is perceived as having greater impact than that same event occurring in the next town, next region, next nation or a continent away. The problem then, is to establish a series of "equal impact contours" for events at successive degrees of removal in space/time. See Figure 12 for the general model.

General questions to explore in this topic include:
Is the "discount rate" of an event's impact the same in the past as in the future?
Do discount rates vary systematically according to group membership?
Is it really perceived likelihood or perceived impact that governs one's assessment of an event?

Particularly for environmentalists, is resistance to construction of nuclear facilities more dependent on the perceived likelihood of an accident, or on the perceived seriousness of an event, given

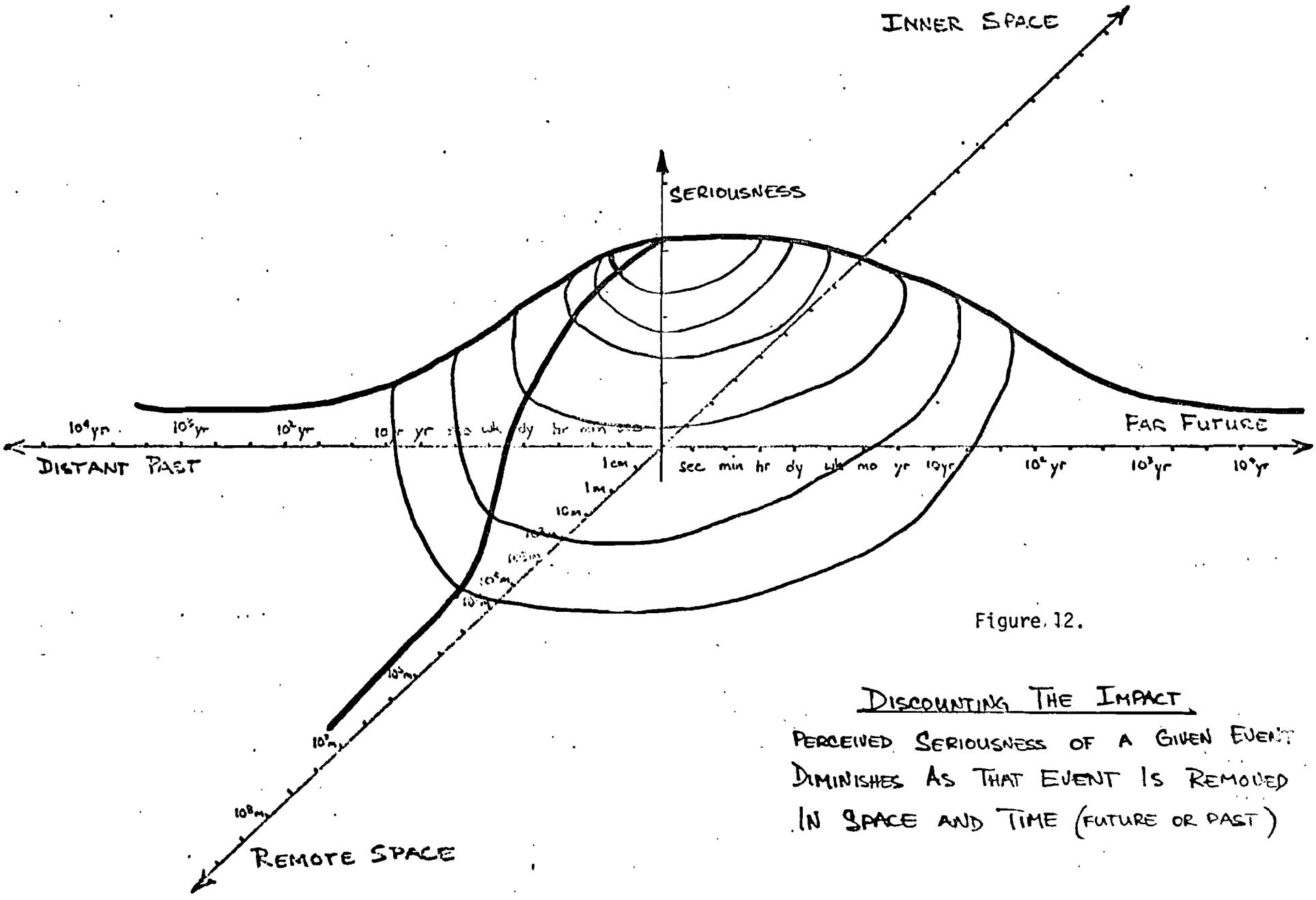


Figure. 12.

DISCOUNTING THE IMPACT

PERCEIVED SERIOUSNESS OF A GIVEN EVENT DIMINISHES AS THAT EVENT IS REMOVED IN SPACE AND TIME (FUTURE OR PAST)

that is occurs?

One should examine such surfaces which describe the assessed risk seriousness for differing special interest groups. Imagine individual surfaces existing for environmental advocates, different surfaces for scientists who develop and implement nuclear installations, others for licensing boards, boards of directors of non-nuclear installations, citizens panels representing land and property developers, mass housing magnates, parents, as energy conservationists, indeed a host of special interest groups can be hypothesized (and perhaps found) in the population whose assessments of impact in need to be considered in order to bound problems of environment, energy and subjective value systems.(Figures 13,14).

Some comparisons of perceived likelihoods of real events compared with known figures should shed some perspective on non-rational assessment styles. The nature of probabilistic events is that their likelihoods are constant over space/time; but one can well imagine that respondents to the survey will not necessarily perceive it so. Understanding this discrepancy may well help in promoting public education news items to reduce non-rational objections.

Documenting the Eventual Findings.

Lucid displays of the findings of such survey procedures is as important as the results themselves. Certainly as a public information vehicle, graphic displays are far more helpful than tabulor or mathematical forms. To give early consideration to how these results might look also helps shape and crystallize the problem specifications and methodologies enroute. This Working Paper is its own instance of the copious use of graphs, tables, etc. as vehicles for thought. Mathematical expressions are more rigorous, but less vernacular modes for expression.

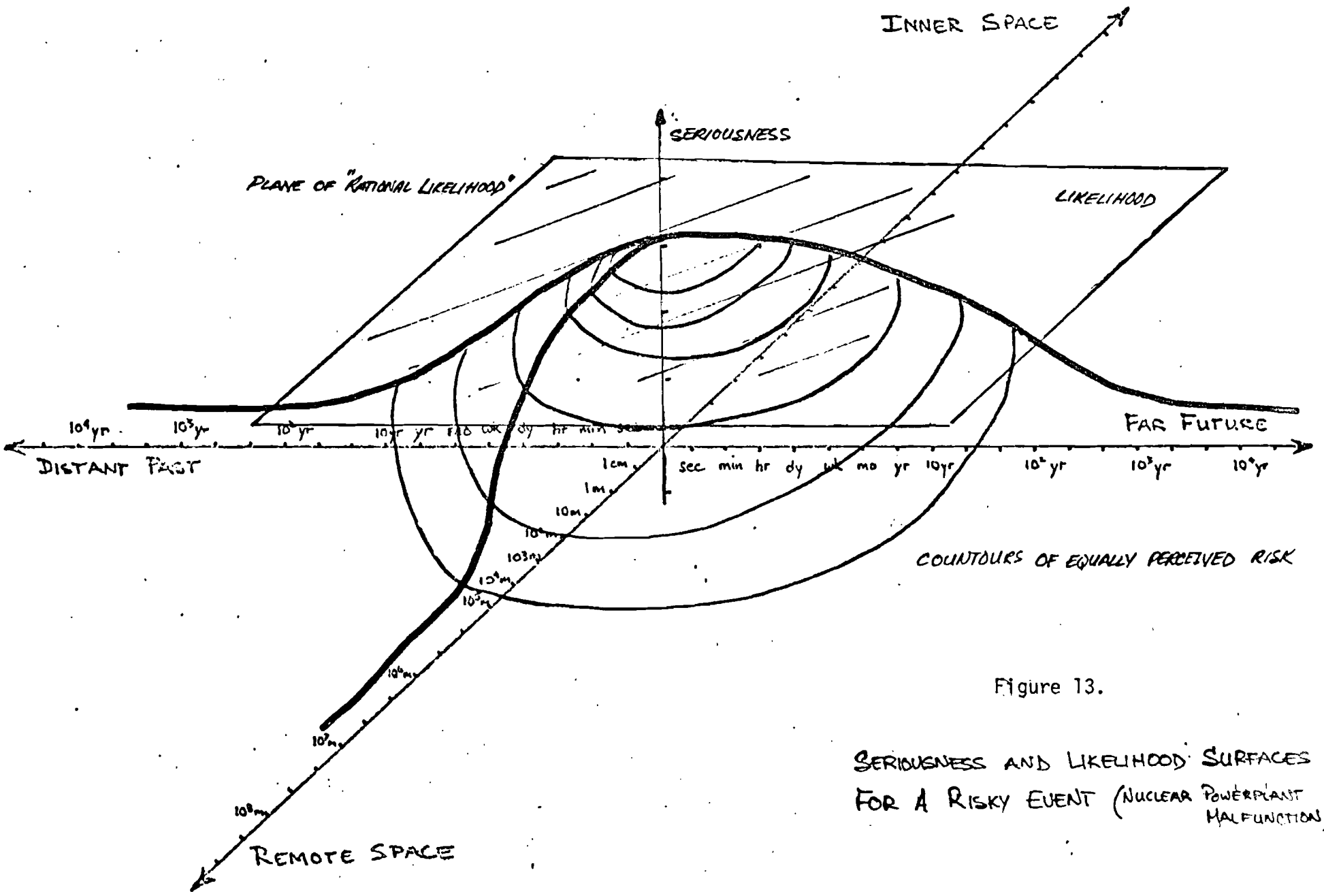


Figure 13.

SERIOUSNESS AND LIKELIHOOD SURFACES FOR A RISKY EVENT (NUCLEAR POWERPLANT MALFUNCTION)

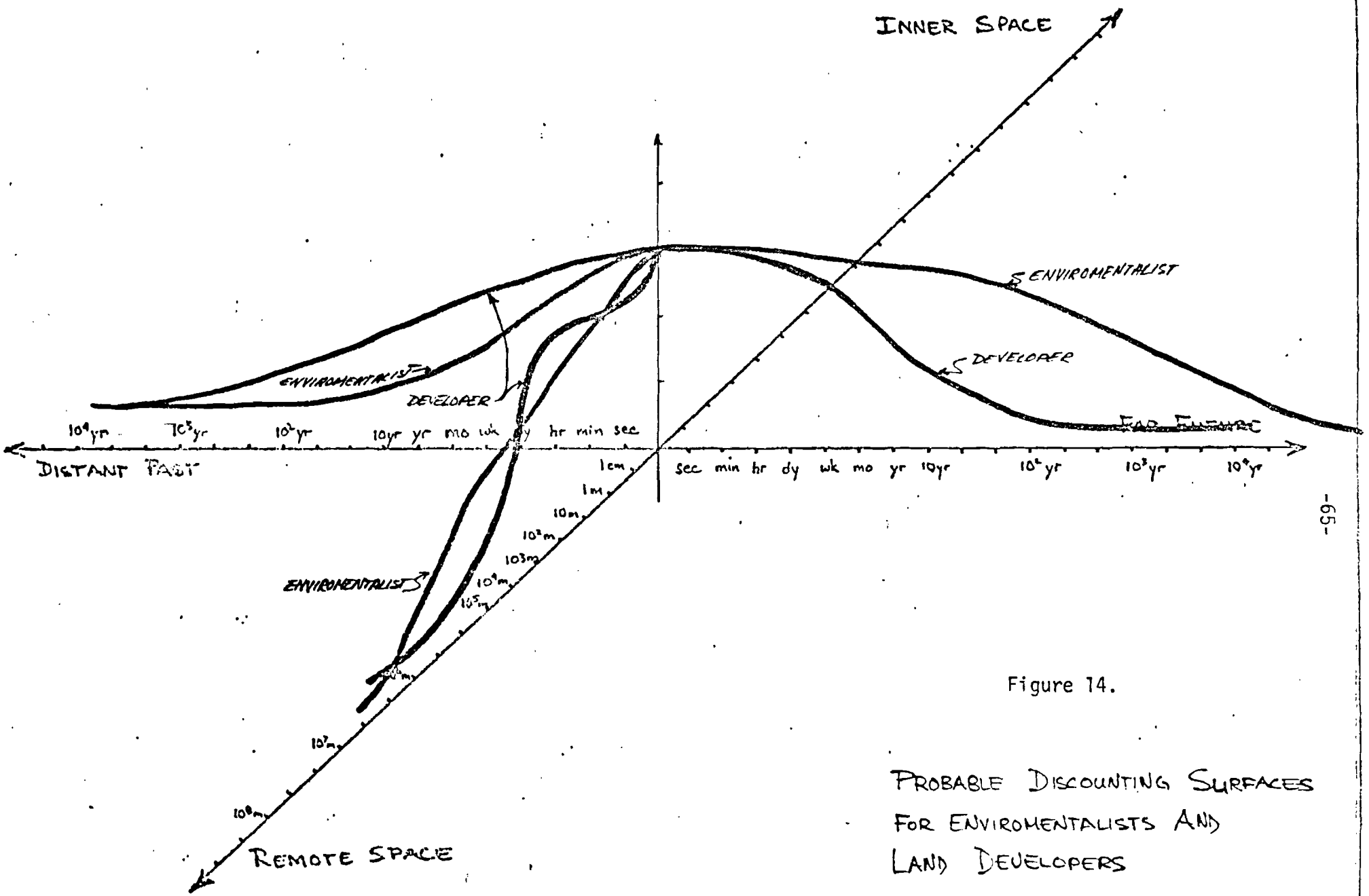
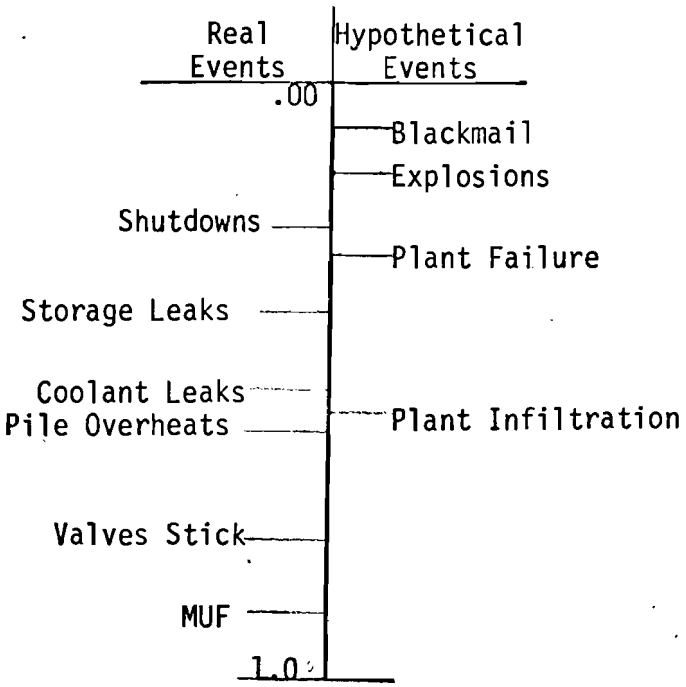


Figure 14.

PROBABLE DISCOUNTING SURFACES FOR ENVIROMENTALISTS AND LAND DEVELOPERS

Recalling Steven's scale of professional prestige on page 31, it's difficult to conceive of a simple yet more badly needed scale than one which might look like this:

Figure X



Comparisons of Perceived Likelihoods for Real and Hypothetical Events

Some earnest time should be spent early in the planning to decide what display techniques will work best at various stages of the survey. Ad-hocracy in this task really should be avoided.

Strategies for Exploring Decision Space

A multi-dimensional decision space can contain myriad strategies for its own exploration. The question is how best to explore this decision space, particularly given the immediate problem facing IIASA. Three complementary strategies are under discussion: Pareto-optimal, Multi-objective, and Magnitude Estimation strategies.

A lengthy discourse on the merits, demerits and implications of each can be spared, but a quick summary is probably in order. Exploring this decision space is rather more simple than has been thought. There are only about three major variables to be considered: 1) Specific Events (impacts, occurrences); 2) Dimensions or criteria; 3) People or groups to do the assessing. Events can be examined against explicit criteria. For example, is this event A more or less risky than event B? Is event C more or less costly than event D? Does event E affect greater or fewer people than event F? Does event G promise greater long-term utility than event H? Will event I require more man hours for its implementation than event J? Where criteria are explicitly stated, the events can be arrayed along the criteria one at a time. Or, secondly, the events may be stated explicitly but the criteria left implicit.

Various special interest groups who may share a common value system may have sharply differing opinions about the importance of a given criterion, hence on the impact of an event on that criterion. It may emerge that such groups differ only as a result of the criterion weight and of the placement of events along each criterion and not uniquely by group

membership. If group membership can be predicted from event assessment plus criterion weighting, then group membership is no longer a necessary distinction.

What can one reasonably require of a technique for exploring a decision space? These requirements fall into about four headings: Nature of the Subjects, Nature of Criteria, Events to be assessed, and Research Convenience (ease and cost of administration). The following table makes some early guesses about the 3 present techniques on each of these requirements. I've played hunches for a good many of the requirements on Pareto and Multi-Objective techniques. Gros, Keeney and helpful new arrivals should correct my errors as a first chore.

Trade-off Parameters for 3 Decision Making Strategies

		Pareto-Optimal	Multi-Objective	Direct Estimation (Stevens Ratio Scaling)
Subjects	Do subjects (S's) require prior training?	yes	yes	no
	How many S's can be surveyed in a single setting?	1	few	any number
	Can competing special interest groups be easily explained?	same as for any single person	yes, a few at a time	yes
	Is face to face contact among opposing viewpoints possible? Necessary?	?	possible, and perhaps necessary	possible, but not necessary
	Do S's manipulate numbers directly?	may do so, or may assess better odds	may do so	may do so, or may also express assessments in terms of value of other events
Criteria/Variables	Are criteria explicit? Implicit?	explicit	implicit	explicit
	Can multiple criteria be explained simultaneously?	no	yes (?)	presumably (not yet documented)
	Can inferences be made from one criterion directly to another without invoking a common criterion such as probability on $ value?	no	no	yes, direct and exact "cross-model matches" can be readily made by S's
	Does "Embeddeness" of the events display itself directly along each criterion	?	probably only as paired comparison	yes
Events	Can many events be surveyed simultaneously?	two at a time	two at a time	yes, possibly more than a hundred
	How are events on any single criterion expressed in terms of each other?	indifference curves	indifference curves	ratios
	Are "Hypothetical" events permitted?	probably	probably	yes
Convenience	Is information about the event conveyed in the surveying?	yes	yes	not usually
	Speed of data collection?	relatively slow, 1 to 1 gaining technique	relatively slow, small group gaming	very fast, data for perhaps 100 can be collected in 30 minutes
	Can opposing viewpoints be surveyed simultaneously?	no	yes	yes
	How good is the "Resolution"?	bounded by zone of indifference curve	bounded by zone of indifference curve	excellent for concrete events, subject to increasing interpersonal variability for remote or hypothetical events

Presumably, a primary concern will be the comparative efficencies of the 3 D-M techniques. The 3 strategies address somewhat different questions, proceed along different routes and invoke different assumptions. Therefore, the costs, payoffs and focuses will vary, too.

In general, I envision Magnitude Estimation as a broad spectrum means for examining the entire domain of energy, costs, environment, risks and benefits. To include it as a complementary strategy ensures that one will not miss the forest for the trees (to coin a cliche).

Again, referring to Steven's own graphs, his interest was to get 'single digit' precision across 3, 4 or 5 orders of magnitude of stimulus intensity, and to link that to as many "subjective octaves" as the respondents chose to produce. In general, the perceptual dimension reduces the span of magnitudes quite considerably.

I worry about the risks of placing too much weight on the results from the pareto-gaming of a handful of specific events which may cover only a tiny fraction of the range of potential event impacts. On the other hand, such close-focused scrutiny will surely produce more exacting data over the spectral width which it <u>does</u> examine. To quote Stevens, "as in any empirical inquiry, much depends on the level of accuracy required."

Examine the foregoing table: Magnitude Estimation can survey large numbers of people, large numbers of events and great ranges of impacts quickly, cheaply, but <u>roughly</u>. The other techniques will produce closer resolution, but at greater cost.

These will be the guiding questions;

How much detail is needed?

How much manpower is available?

Does respondent's experience with our technique affect his performance on other techniques?

What special training of respondents and of administration is required? Does this affect performance?

Does face to face gaming among the groups produce better (more realistic) results?

To answer these questions (particularly for future undertakings) presupposes that some careful in-house record keeping occur. Tooling-up costs are as much a part of the real costs as are field expenses, and should not be forgotten.

The following section suggests a possible framework to answer most of these questions, but the questions should be elaborated before any real efforts are expended.

Procedure for a Test Case

Since the question of Public Values is guaranteed to emerge repeatedly as costs for energy (or commodities or services) escalate, it would be convenient to have a ready strategy for assessing the extent of public valuation which attaches to the heretofore "non-costable" components of goods - whether they be tangibles or intangibles like energy or services.

This subsection sets forth a framework, a task list and a time budget which would be typical of any effort to capture the extent of public awareness, willigness and support for a new facility. That there are many more specifics, unknowns and variables than those elaborated here, is obvious. Nevertheless, as unknowns became knowns and as unwitting issues emerge, it is a simple task to fit them in their appropriate context within this framework and proceed as shown.

In this "Test Case" I have assumed that the "Case" is a undertaking with region-wide,or even national impact; that public opinion would run strong, that questions of legality and right of public good would be pitted against the antagonists of such a project. I had a nuclear power plant or a nuclear fuel reprocessing plant specifically in mind, and either would exist at the approximate scale which this "Test Case" addresses. I further assumed that a set of Public Hearings and a Court Ruling would be useful vehicles to underscore the seriousness of the Case in the public mind. Conceivably the Court Ruling could be

on several potential wordings ranging from: "that construction for a nuclear fuel reprocessing plant at location X proceed" to "that permission be granted to proceed with selection of a site for a nuclear fuel reprocessing plant to be built in the future". Thus the fact of a court ruling need not be a necessarily incendiary issue.

A list of some 14 "Tasks" encompasses the major issue set which must be pre-specified before an undertaking of the magnitude envisioned could be begun. Surely, this list is likely to expand - - perhaps as the specificity and site location of any such project becomes more immediate and definable. There is the further assumption that some clear strategy is previously at hand for systematizing the issues to be presented to the public for assessment; hence, the accompanying Event Generating Machine.

Clearly, it was my personal intent that all 3 of the strategies of the previous sections be employed in examining a test case in order to test the range of convenience, the resolutions, administration ease and cost of each. The "Event Generating Matrix" is an indespensible tool for rationalizing (well in advance) just which events and which criteria shall be examined by the public. Persons representing the public section will doubtless want to elaborate new criteria and additional specific events, but to rely solely on these spokesmen to suggest the full range of events and criteria is to court both methodological disaster and public ridicule.

This list of Tasks is the general framework which will guide the planning, surveying and followup phases. There is no man-hour complement yet attached to the time budget. The staffing ratio is too dependant on local conditions to be intelligently specified (or even suggested) at this distance.

14 Major Tasks Forcing a Test Case

1. Detailed Planning
2. Generating an Extensive Events List
   Location Selection Issues
3. Possible Locations for the Plant
4. Locations for Collecting Surveying Information
5. Identifying the Special Interest Groups
6. Promoting Public Awareness About Survey (Media)
7. Soliciting Representations from
8. Holding Formal Legal Hearings
9. Meeting with the Special Interest Groups (3 Sessions Each)
10. Administering the Survey (3 Techniques)
11. Analyzing the "Events Assessments" and Value Judgements
12. Reporting Values to the Public via Media
13. Obtaining Court Ruling to Proceed (toward explicit goal)
14. Analyzing the Relative Merits of the 3 Alternate Strategies

Assume, then, that a major project is to be constructed - - an installation as large and as controversial as a reprocessing plant, or a large nuclear power plant. What issues must be examined; in what order; how does the public sector become aware, involved; how is the public mind best assessed. Here is a set of task in approximate order and detail to proceed:

Detailed Planning

Clearly, in an undertaking of this size: "If you don't know where you're headed, you're not likely to get there." Early, detailed and thorough planning are essential. There will be ample time and opportunity to change details, strategies, personnel, responsibilities

enroute if such should become necessary. This subsection itself is perhaps the first element in the planning, but local assignment of responsibilities, tasks and overall directions is expedient. 3 or 4 people with suitable assistance can accomplish the major detailing, scheduling and pre-contacting of essential persons in 3 to 4 weeks, once the exact scope of the 'Test Case' is known.

Generating a List of Events

Once agreements are finalized on whether a reprocessing plant or some alternate facility is to be the test case, the team can begin to produce an exhaustive list of EVENTS. It has been argued previously that for IIASA to undertake a very extensive survey of "Secondary Attitudes" or "Supportive Opinions" is inefficient. If the issue at hand is Public Support for Nuclear Reprocessing Plant, then ask about Processing Plants and the events which surround the construction and operation of one - - not about ". . . isms" (Urbanism, Environmentalism) which might be related to popular support of such a plant.

This list of EVENTS will obviously have two major components:

1) Known, demonstrable events associated with construction and operation: waste heat, MUF, spurious radiation, size of plant buildings, additional committed land areas, public safety, public access, structural esthetics, risks of construction and operation, storage of radioactive input and output materials.
2) Hypothetical events: systematic division of MUF, equipment breakdown, leakage, explosions, international blackmail, etc.

As argued previously, to validly assess any event, it must be placed in context of many other events equally germane to one single criterion (at a time). Most often, the "pool" will have to be supplied to the assessors, together with suitable instruction on how to make these assessments. They may well want (and demand) to supply additional events ("worries") and this should be welcomed, but the intelligent assessment of even these voluntary events itself depends on fitting them into some available context. Hence the need for a standard, IIASA supplied list.

Records of similar plants, expert knowledge and "brainstorming" will yield several hundred such events. The Matrix affords a classification scheme, shows where empty or thinly covered cells remain. U.S. AEC records for last year (1973) showed 861 "abnormal" occurences. They would be a good starting point. (See Appendix)

Potential Location Selection

Two kinds of location issues emerge as needing early resolutions.

1) which regions of the country are reasonable candidates of location of such a plant; and,

2) from how wide a geo-political span should public representation be encouraged.

1) Potential Regional Location

This level of analysis is probably too general to address the siting question (where "site" is understood to be the several square kilometres on which the actual structure will be constructed; and "location" is used to mean the region, Bundesland, etc.) There is a very real "Zone of Intrusion" beyond which I.I.A.S.A. will have gone too far if the residents of the region have not yet been included in the deliberations -- whether they be "Real" Test Case of "Hypothetical" Test Case. Deliberations are in order, however, to determine whether a Reprocessing Plant will more likely be in the North or South of Germany, perhaps even whether it should be Rhine based-plant, or North Sea. Beyond this, I.I.A.S.A. (as I.I.A.S.A.) best be silent.

2) Survey Location

Some early decision should be struck whether that public sector to participate in the survey will be located near the potential Plant region; or whether the pioneering aspect of such a plant and of such a survey dictates that a multi-region or even national survey is in order.

The more vocal groups will make themselves heard irrespective of where potential sites may be; but these are the groups whom one could not miss anyway. Less strident groups and individuals may have to be actively solicited after the regional locations are established.

Identifying Groups

Groups will be identified in several ways.

1) Some groups (and their spokesmen) are already known. These should be listed immediately, together with a brief characterization of their platform.
2) Other groups would identify themselves if a region-wide or nation-wide announcement were to be made calling for "position papers" on a proposed plant.
3) As soon as a region (city) and a firm proposal for such a plant were announced, still others will emerge.
4) Public hearings to "receive representations" on such a proposed plant and location will produce still more.
5) Newspaper and Television items announcing proposed plants and locations will produce further contributions.
6) Court hearings will weed the serious contenders from the Sunday Declamators.

The difficulty, of course, is that the objectors are usually more shrill than the proponents. Thus, special attention will need to be paid to finding public representatives who favour construction (for reasons other than speculative interest in the land) whose numbers match the opponents. Professional scientists, university students and futurists may have to be deliberately sought.

How many such representatives are needed? There is no hard and fast answer. Factors which govern the decision are: How many people in each group make themselves known? What exactly do you want to know about differing viewpoints? How big a difference between averages of two (or

more) groups' opinions do you believe would be noteworthy? How many people would it take to make such a difference statistically useful? The easy rule of thumb is: "If you can get 30 in each group, you have a fair chance of demonstrating any sizable differences that do exist". I would be personally uncomfortable if a formally recognized platform had fewer than 10 advocates.

Public Awareness (Newspapers, TV, etc.)

The media services need to be given special attention in the early planning since they will serve on several fronts: 1) to awaken public awareness to the idea that (in the short run, at least) environmental preservation is partially at variance with public demand for more and more energy, 2) to publicize that these "social and public values" are being systematically examined, 3) to direct people where to make written or personal submissions, 4) (after the survey) to report what people's general value structures are.

Solicited Representation

Many organized groups and private citizens will have closely reasoned arguments which outline and support their positions. These need to be collected and carefully analyzed for issue, content, evidence and strategy before approaching the public. Each group's general argument as well as approximate event configuration can be inferred from these representations, written or otherwise. Besides their informational value, soliciting such representations insulates the Test Case against criticism that "Nobody asked us!"

Legal Hearings

Legal Hearings are a more formal extension of the solicited representation medium, but with the additional double advantage that they are formalized in a legal and publicity recorded format. They will produce face-to-face confrontation, thus I.I.A.S.A. investigations are early exposure to the"real world" positions, tactics and decision space of the legal forum.

Meeting with the Groups

The eight preceding tasks will bring forward those special interest parties who have particular reason for wanting their position heard, aired and considered. As each group or strong spokesman makes his existence known, the I.I.A.S.A. team must introduce themselves, state their interest in a"special opinion survey" and collect names and contact points (addresses, phone numbers, etc). Team members need to have a sufficient scope of the politics and social ordering to determine at the first contact whether the person is part of a group, which one, etc. By far, the most skillful way to do this is to present the person a small pre-printed file card asking for name, address, phone, group membership, approximate platform and willingness to participate in a future survey. This way, even those declining to participate are allowed the opportunity to contribute to the overall pattern of public sentiment.(See Appendix).

Survey Administration

The time budget following, shows three contacts with each of the group classes (six are shown, but that is a purely artitrary number).Whether three, more or fewer sessions are necessary is an issue for local decision. I had envisioned one session of orientation, familiarization, introduction, pretraining where necessary; a second session of actual assessment via one or more of the three strategies, and a final, third session of assessment following feedback of his own and other groups assessment. In essence, it forms a two iteration Delphi Technique with or without the "justification stage" of the pure Delphi format.

Whether groups meet separately, jointly with other groups, or even face-to-face gaming with members of other groups will be decided by the particular assessment technique.

To fail to monitor the effect of prior exposure to the assessment procedure, passage of time following the procedure and the effects of information feedback on subsequent assessments would be a serious and unnecessary loss to this Test Case.

G. Baecher, J. Gros and R. Keeny will add modifications, idiosyncracies, and fine tuning to this rough framework, all of which should be debated for inclusion to afford the "fine grained, close resolution" look at prevailing public valuation of the several criteria.

Event (and Criterion) Analysis.

Obviously this phase is the heart of the entire undertaking. Data regarding individual events and their respective placements along several criteria is the Central obejctive of the project. But including event analysis at this particular point in the task sequence has a particular rationale.

If the results of informational feedback are to be tested to see if the availability of such assessment information alters the subsequent assessment of the same events on the same criteria (i.e. does education make a difference?) then clearly the analysis must proceed simultaneously with the survey administration or lagged only slighly behind it.

The nature of Steven's Magnitude Estimation technique is such that anyone with a Hewlitt-Packard Pocket Calculator, a piece of graph paper and a pencil can provide feedback information to about 20 or 30 people for 8 or 10 criteria in about a half-hour.

Whether Pareto-Optimal and Multi-Objective Assessment Techniques have such brief turn-around times can be determined quickly at I.I.A.S.A.

Turn-around time for Pareto-Optimal _____( ).

Turn-around time for Multi-Objective _____( ).

In any case, the version of the time budget shown here allows for ample time between meeting sessions, so that no matter which strategies are employed, the results of the previous session can be displayed and discussed prior to the second assessment.

Therefore, the time budget shows Event Analysis begining almost immediately after the first groups have been surveyed. As shown, the information from all six (hypothetical) groups can be collected and analyzed in time to display to the next session of even the first group.

Event analysis continues beyond the last sessions with each group in order to summarize the information from both session from all groups to forward to the media, legal hearings and the hearings for the court ruling.

Reporting Values to the Public

It would be a misuse of I.I.A.S.A.'s position if information of this sort on an issue of this gravity were not forwarded quickly to the public. Indeed, the speedy presentation of the subjective and hypothetical event values are themselves means to further public awareness of true costs of energy resources. Thus, as quickly as possible, the event and criterion analysis should be rendered in non-technical language and presented to the public by means of News Releases, Television Spots, etc. Results in the technical and scientific journals need less explicit advocacy. They will appear perforce.

Obtaining Ruling

If the decision to apply for a court ruling on some aspect of the project is implemented, then the legal hearings as well as solicited representation become direct and non-trivial input into the court records of the deliberation. These documents together with the actual data of the assessments of events afford the opportunity to compare the "positive" decision-making inputs with their "predictive" utility in anticipating the court's decision. Either outcome will be instructive in evaluating the utility of formalized decision-making procedures.

The ruling itself, if positive, clears the way for the next stages of planning, and construction; if negative, affords a test of formalized decision techniques with the 'real' geometry of the legal decision-space.

Strategy Analysis

A useful spin-off of the entire 'test case' is its insight into the relative advantages of three (or more) decision-making strategies. How do they compare in cost, ease, range of convenience, resolution of data, interpretability, mathematical tractability etc. (See Baecher's Notes, August 20, 1974)? Is the "broad brush-stroke" or "fine tuning" hypothesis confirmed? How much pre-training is needed to enable people to perform the assessments? etc. etc.

Clearly, these questions of strategy, cost yield, and convenience can be answered at a more leisurely pace than can the event analyses; still their value is equally useful in framing new ways for future "truth surveys".

Logically, some of these tasks must fully precede others; some can partially overlap each other, some (such as event analysis and survey administration) must run concurrently as that information can be continuously updated; while still others bear no logical contingencies to each other. Figure 13 following, shows the sequential ordering of each of the fourteen Tables with respect to each other. The Figure reads downward, column by column. The entry indicates whether the column task must (P) - precede the row task, (O) - overlaps with its, (S) - simultaneously occurs, (I) - is identical with it or, (F) - follow the row task.

Figure 14 is the Time Budget for the 14 tasks. Forty-four weeks have been allowed for leisurely accomplishment of all tasks. Surely, some tasks can be compressed into a fraction of the time shown here, if sufficient manpower is available. Other tasks, particularly Meeting with Groups, Survey Administration and Event Analysis will keep 10 or 12 people quite busy in arranging Public Meetings, fielding the media, preparing information releases to the public, analyzing and charting each sessions' results to feedback into the next session. If public turnout is large, even more personnel may be necessary.

Note: The I.I.A.S.A. planning team will have to exercise particular wisdom in the choice of personnel to conduct the Hearings and Public Meetings and to Administer the Surveys. Probably the Laboratory staff recently under Professor Haefele's direction is appropriate. For public assessment of this sort, knowledge of the local conditions, vernacular communication skills and extreme language fluency are demanded, as well as considerable ease in facing several people -- sometimes even hostile ones.

Depending on I.I.A.S.A's perception of available time, an exercise in com-

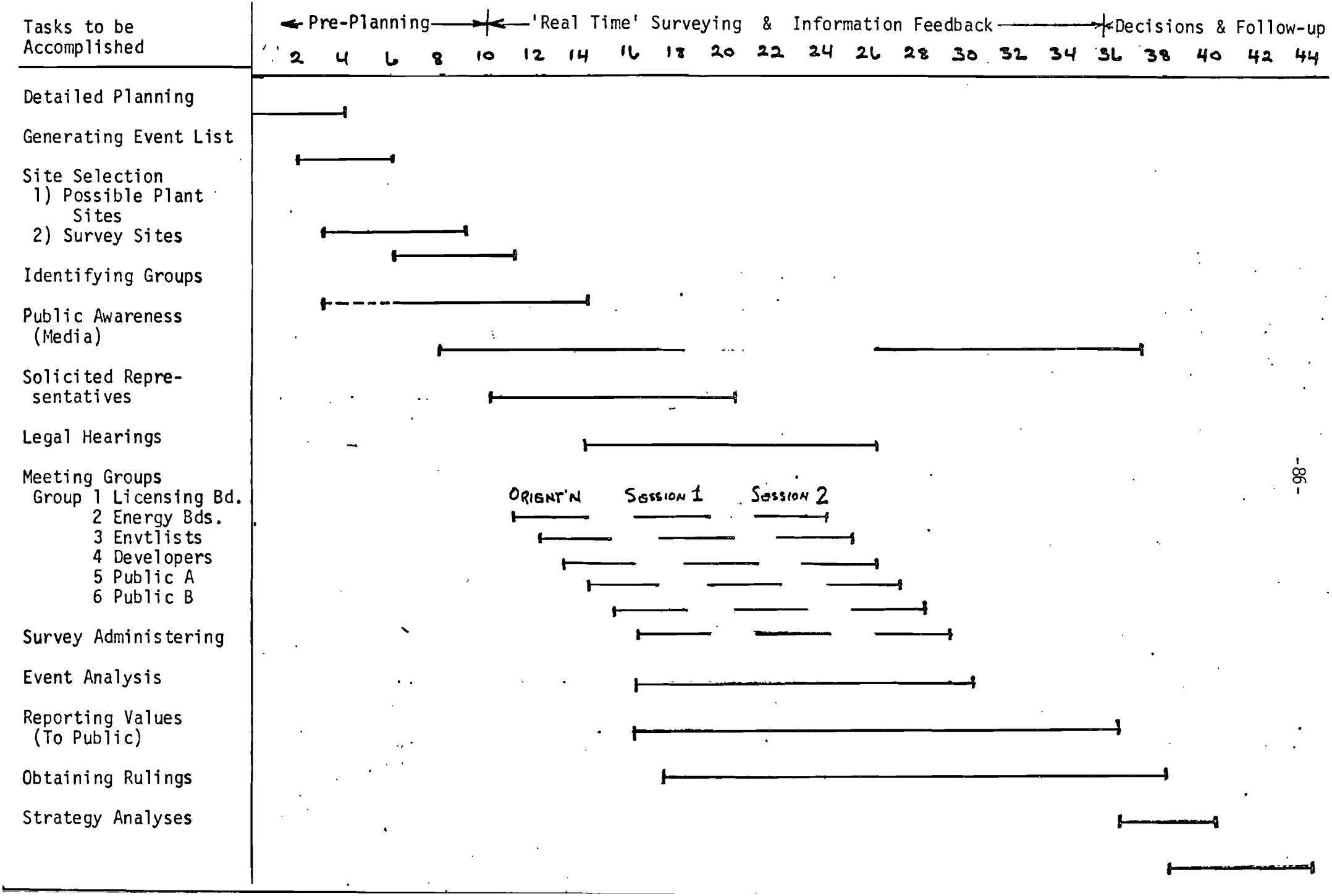
INTERRELATIONSHIPS AND SEQUENCING - 14 TASKS

P - Column Task must Precede Row Task
O - Overlap in Task Sequencing
S - Simultaneous Information Exchange
I - Identity

Tasks to be Accomplished	Pre-Planning				Surveying										Decisions		
	DP	EL	PPS	PSS	IG	PA	SR	LH	1	2	3	4	5	6	SA	EA	RV	OR	SA
Detailed Planning	I																		
Generating Event List	P	I																	
Site Selection																			
1) Possible Plant Sites	O	P	I																
2) Survey Sites	O	P	O	I															
Identifying Groups	O	P	O	O	I														
Public Awareness (Media)	P	P	O	P	O	I													
Solicited Representatives	P	O	O	O	O	O	I												
Legal Hearings	P	P	P	P	P	O	P	I											
Meeting Groups																			
Group 1 Licensing Bd.	P	O	P	P	O	O	O	O	I										
2 Energy Bds.	P	O	P	P	O	O	O	O	O	I									
3 Envtlists	P	O	P	P	O	O	O	O	O	O	I								
4 Developers	P	O	P	P	O	O	O	O	O	O	O	I							
5 Public A	P	O	P	P	O	O	O	O	O	O	O	O	I						
6 Public B	P	O	P	P	O	O	O	O	O	O	O	O	O	I					
Survey Administering	P	P	P	P	P	P	P	O	S	S	S	S	S	S	I				
Event Analysis	P	P	P	P	P	O	P	O	S	S	S	S	S	S	S	I			
Reporting Values (To Public)	P	P	P	P	P	S	P	S	S	S	S	S	S	S	S	S	I		
Obtaining Rulings	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	I	
Strategy Analyses	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	O	I

TIME DEVOTED TO EACH STAGE (WEEKS)

(Full Scale Survey can be done in 10 months easily)

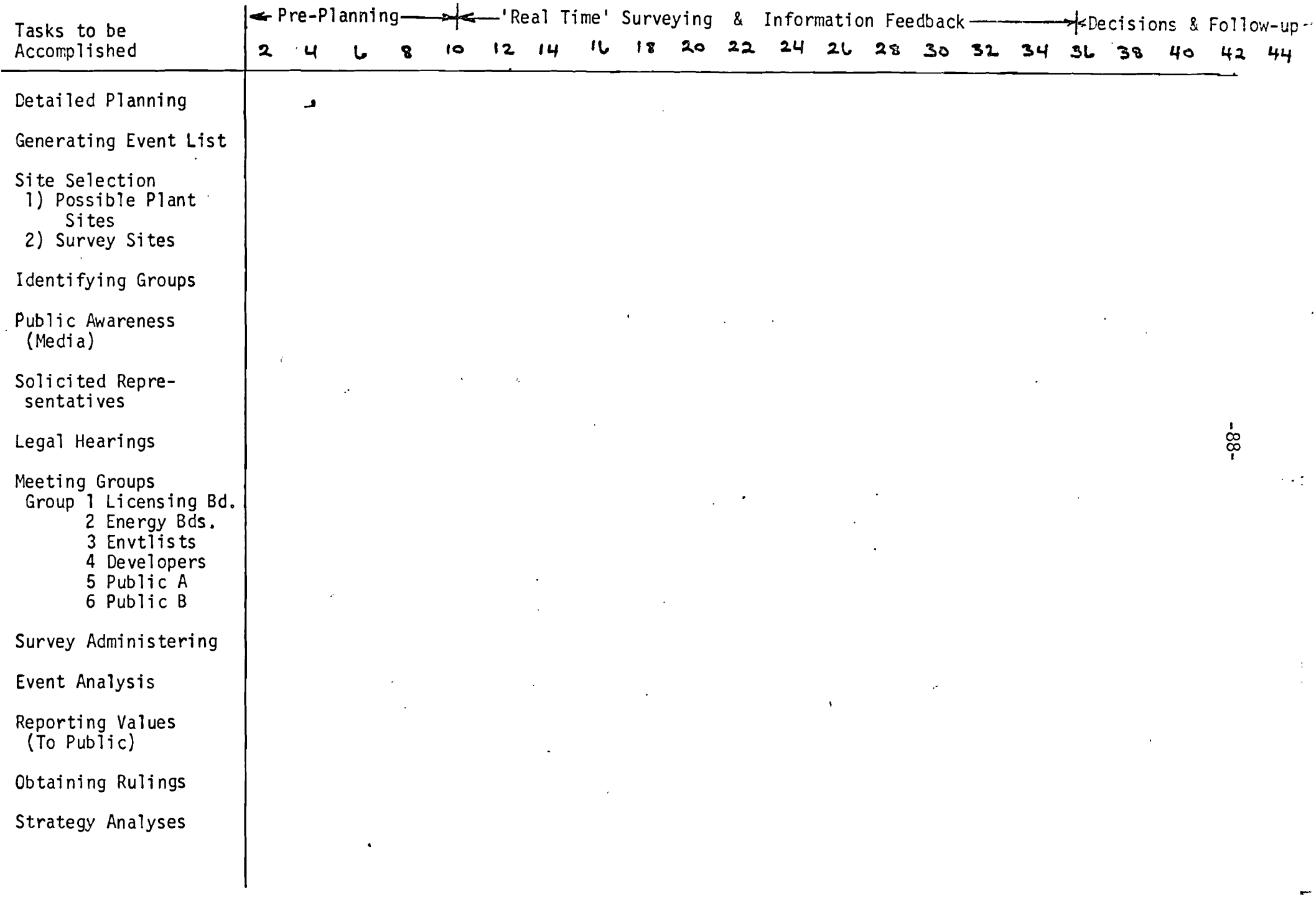


pression may well be warranted. Figure 15 is the same time budget with the time scale and task duration removed. Team members should (independently and in concert) decide just how far the Time Budget can be compressed without losing careful planning and attention to detail. A useful first step would be to delete items that now appear unnecessary (or redundant), but to insert tasks that have emerged as essential, given the further study since September. (One critical planning item would seen to me to be the pre-programming of the Events Analysis for the three techniques, either as a computerized package, or as a set of "crib sheets" which facilitate the calculation, but also demand prior clear decisions of what to do with the data. Display techniques, graphs, charts, etc., for showing the results of the analyses to the second assessment session and to the public should be pre-planned well in advance of the field work. Some phases of the survey must resemble well-rehearsed theatre more than hard-nosed science).

The final Task List and Time Budget may bear little resemblence to these items shown here, but as a point of entry into a rational sequencing of survey planning, these notes should prove useful. (Would that we had done it for the Vancouver Urban Futures Project).

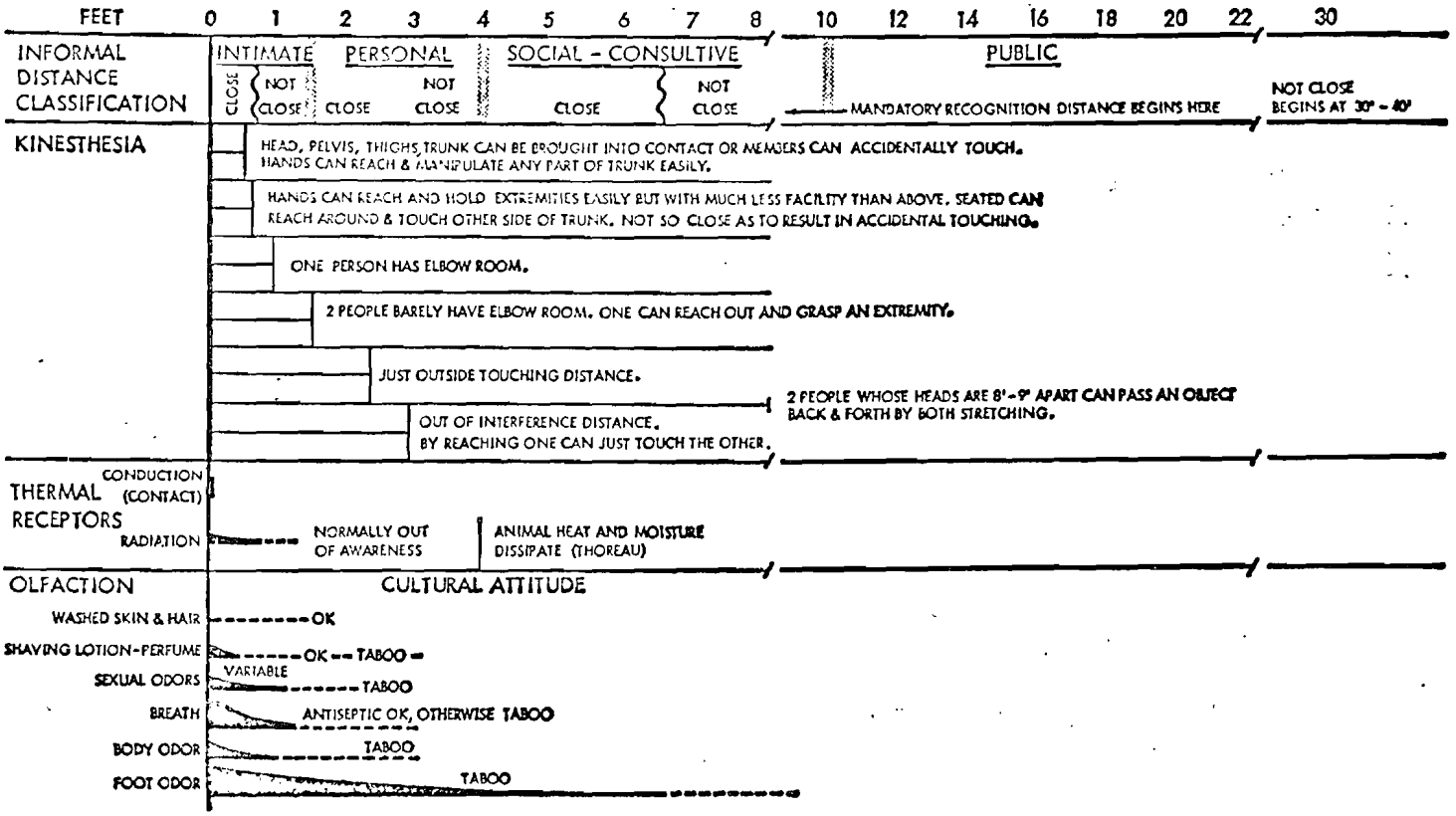
TIME DEVOTED TO EACH STAGE (WEEKS)

(Full Scale Survey can be done in 10 months easily)



APPENDIX

CHART SHOWING INTERPLAY OF THE DISTANT AND IMMEDIATE RECEPTORS IN PROXEMIC PERCEPTION



PRESTIGE OF TWENTY OCCUPATIONS
MAGNITUDE SCALING VS. CONFUSION SCALING
ASSESSED BY THIRTY UNIVERSITY STUDENTS
NOVEMBER, 1974 U.B.C.
The advantages of being able to report that profession A is 4 times more prestigous than profession B are perhaps not immediately obvious until one wonders about the corresponding dimentions that might be associated with prestige. Do the various professions pay a corresponding amount? Do they require equal ratios of educational preparation? It is a hopeful development to say (for example) that a Farmers job is 3X a prestigous as a Farmhand's; or an Engineer is 7X; or a Physician is 10X.
PHYSICIAN (12)
UNIVERSITY PROFESSOR (17)
ARCHITECT (3)
ENGINEER (8)
ARTIST (5)
HIGH-SCHOOL TEACHER (4)
PHOTOGRAPHER (10)
MANUFACTURER (13)
FORESTRY OFFICER (14)
ELEMENTARY SCHOOL TEACHER (18)
FARMER (7)
SALESMAN (11)
SAILOR (15)
PAINTER (16)
HOUSEWIFE (6)
BUS CONDUCTOR (19)
HAIRDRESSER (9)
FARMHAND (2)
FACTORY WORKER (20)
CASHIER (1)
TEN-POINT (CONFUSION) SCALE
Magnitude Estimation Scale

File Card for Respondents

Name

Address Telephone

Are you here representing the general public,
or an organized Group _____(public) _____(group)
Name of the Group _____
Do you personally favour or oppose further planning
for the proposed Plant? _____(favour) _____(oppose)
Major Reasons _____
Would you attend Public Meetings on the _____,
_____, and _____ to discuss these issues? __(Y) __(N)
Will you or your Group make a written representation?
_____(I will) _____(Group will) _____(Neither)

A deadly quandry

Safe home needed for plutonium

WASHINGTON (CDN) — Atomic energy officials are in a quandary over how to dispose safely of Plutonium 239, the extremely radioactive and deadly substance left over from nuclear power production.

Because plutonium is 20,000 times more toxic than cobra venom, and because inhalation of one milligram (the size of a pinhead) is enough to kill a person within a couple of days, nuclear scientists consider its disposal to be more important than any other public safety issue in the U.S.

The Atomic Energy Commission is holding a public hearing in Salt Lake City Thursday to consider potential sites for disposing of plutonium and other highly radioactive wastes. At a hearing two weeks ago in Washington, environmentalists urged the AEC to suspend all nuclear power development until a solution to the waste disposal problem is found.

The volume of plutonium is not large. The difficulty is that it must be isolated from the earth's living environment for up to a million years.

Various waste disposal schemes have been proposed — removal to outer space, burial in the icecaps of Antarctic, and disposal in deep subterranean cavities created by nuclear explosions, to name a few. For the time being, the AEC is weighing two alternatives — burying the wastes deep underground in natural salt deposits, or encapsulating them in surface storage facilities.

The salt deposits are in stable areas where earthquakes are very infrequent. But disposal in salt means the wastes would become irretrievable after a few years. Heat from plutonium or other highly radioactive substances would cause the salt to flow plastically around the steel canisters containing the waste, sealing them off completely. The canisters themselves would be eaten away by the salt within a short time. Consequently, the chance of ever retrieving the wastes would be lost, and with it the ability to cope with unforeseen problems.

Holding the wastes in surface facilities would permit retrieval. They could be moved to other sites if endangered. But the price of this approach is eternal vigilance over the waste stores, ensuring their immunity from floods, earthquakes and wars.

The AEC has decided to buy time by building a temporary storage facility where the wastes could be kept until a long-term solution is found. The AEC has proposed locating the storage facility at one of three reactor test sites — in Nevada, Idaho or Washington.

One disposal technique being considered by the AEC involves solidifying the unburned plutonium and uranium wastes from nuclear fuel reprocessing plants, and encapsulating them in stainless steel canisters. About 10 canisters, each one foot in diameter and 10 feet long with a capacity of about six cubic feet, would contain the wastes produced each year by an average-sized, 1,000-megawatt nuclear power plant, according to AEC estimates.

These canisters would be encased in individual concrete shields, and placed within a well-guarded AEC installation. One reason for the guards is that plutonium can be used to construct a nuclear weapon.

Currently, most of the plutonium is being stored temporarily at a commercial reprocessing plant in upstate New York.

Under present regulations, these wastes must be solidified within five years and be shipped to an AEC installation for disposal within 10 years.

So far only small amounts of plutonium have been produced, and the bulk of high-level radioactive wastes has been reduced to a form suitable for long-term storage. But plutonium is the key fuel for the high-speed breeder reactor now in the development stage, and its widespread use lies ahead.

Despite a great deal of effort on the part of the AEC and the nuclear power industry to reassure the public that no safety hazards exist, many people remain unconvinced. While no serious accident has occurred in the nuclear power industry so far, a recent AEC review shows that during 1973 there were 861 "abnormal occurrences" at operating nuclear plants, including several major leaks of radioactive wastes from storage facilities.

The Union of Concerned Scientists, which opposes nuclear development, has said there is no adequate safety program in existence to deal with the problems of plutonium wastes.

Some scientists like George Wald, Nobel Prize winner and professor of biology at Harvard University, believe there is no solution to the waste disposal problem. "Where," he asked not long ago, "is there a place on earth where we can guarantee geographic, geological and political stability for millions of years?"

Notes and References for A Metric for the Social Consensus

References and Notes

1. L. L. Thurstone, "Psychophysical analysis," *Am. J. Psychol.* 38, 368–389 (1927); *The Measurement of Values* (Univ. of Chicago Press, Chicago, Ill., 1959).
2. S. S. Stevens and E. H. Galanter, "Ratio scales and category scales for a dozen perceptual continua," *J. Exptl. Psychol.* 54, 377–411 (1957).
3. S. S. Stevens, "On the psychophysical law," *Psychol. Rev.* 64, 153–181 (1957).
4. ———, "Problems and methods of psychophysics," *Psychol. Bull.* 55, 177–196 (1958).
5. ———, "The surprising simplicity of sensory metrics," *Am. Psychologist* 17, 29–39 (1962).
6. ——— and J. Volkmann, "The relation of pitch to frequency: A revised scale," *Am. J. Psychol.* 53, 329–353 (1940).
7. F. M. Urban, *The Application of Statistical Methods to the Problems of Psychophysics* (Psychological Clinic Press, Philadelphia, 1908).
8. S. S. Stevens, "On the uses of poikilitic functions," in *Stimulus Generalization*, D. J. Mostofsky, Ed. (Stanford Univ. Press, Stanford, Calif., 1965), pp. 24–29.
9. E. Galanter and S. Messick, "The relation between category and magnitude scales of loudness," *Psychol. Rev.* 68, 363–372 (1961).
10. L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (Univ. of Chicago Press, Chicago, Ill., 1929).
11. B. Finnie and R. D. Luce, "Magnitude-estimation, Pair-comparison and Successive-interval Scales of Attitude Items," *Dept. Psychol. Univ. Penn. Mem. MP-9* (1960). Finnie pursued the problem much further in a Ph.D. thesis entitled "An empirical comparison of magnitude and category scaling procedures applied to nonphysical stimuli," Harvard University, 1965. He compared magnitude and confusion scales for several kinds of nonmetric stimuli: attitude statements about the church, adjectives, adverbs, and occupational prestige, and statements expressing various degrees of cynicism. He also scaled the inverse continuum—for example, the degree of unfavorableness expressed by a statement. This work provides impressive new evidence to support the generality of the relation depicted in Fig. 3.
12. G. Ekman, "Discriminal sensitivity on the subjective continuum," *Acta Psychol.* 12, 233–243 (1956); "Weber's law and related functions," *J. Psychol.* 47, 343–352 (1959).
13. R. S. Harper and S. S. Stevens, "A psychological scale of weight and a formula for its derivation," *Am. J. Psychol.* 61, 343–351 (1948).
14. G. Ekman and T. Künnapas, "Subjective Dispersion and the Weber Fraction," *Psychol. Lab. Univ. Stockholm Rept. No. 41* (1957); M. Björkman, "Some Relationships between Psychophysical Parameters," *Psychol. Lab. Univ. Stockholm Rept. No. 65* (1958); "Variability Data and Direct Quantitative Judgment for Scaling Subjective Magnitude," *Psychol. Lab. Univ. Stockholm Rept. No. 78* (1960); H. Eisler, "Empirical test of a model relating magnitude and category scales," *Scand. J. Psychol.* 3, 88–96 (1962).
15. S. S. Stevens, "Review of L. L. Thurstone: The measurement of values," *Contemp. Psychol.* 4, 388–389 (1959).
16. T. Indow ["An example of motivation research applied to product design"], published in Japanese in *Chosa to gijutsu No. 102* (1961), pp. 45–60.
17. G. Ekman and T. Künnapas, "Note on direct and indirect scaling methods," *Psychol. Rept.* 6, 174 (1960); "Measurement of aesthetic value of 'direct' and 'indirect' methods," *Scand. J. Psychol.* 3, 33–39 (1962).
18. G. Ekman, "Two generalized ratio scaling methods," *J. Psychol.* 45, 287–295 (1958).
19. ——— and T. Künnapas, "Scales of aesthetic value," *Percept. Motor Skills* 14, 19–26 (1962).
20. S. D. Koh, "Scaling musical preferences," *J. Exptl. Psychol.* 70, 79–82 (1965).
21. G. Ekman and T. Künnapas, "A further study of direct and indirect scaling methods," *Scand. J. Psychol.* 4, 77–80 (1963).
22. S. I. Perloe, "The relation between category-rating and magnitude-estimation judgments of occupational prestige," *Am. J. Psychol.* 76, 395–403 (1963).
23. S. S. Stevens and M. Guirao, "Subjective scaling of length and area and the matching of length to loudness and brightness," *J. Exptl. Psychol.* 66, 177–186 (1963).
24. S. S. Stevens, "The measurement of loudness," *J. Acoust. Soc. Am.* 27, 815–829 (1955).
25. T. Künnapas and I. Wikström, "Measurement of Occupational Preferences. A Comparison of Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 156* (1963).
26. T. Engen and D. H. McBurney, "Magnitude and category scales of the pleasantness of odors," *J. Exptl. Psychol.* 68, 435–440 (1964).
27. G. Ekman and T. Künnapas, "Scales of conservatism," *Percept. Motor Skills* 16, 329–334 (1963).
28. T. Künnapas and M. Sillén, "Measurement of 'Political' Preferences. A Comparison of Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 172* (1964).
29. S. S. Stevens, M. Guirao, A. W. Slawson, "Loudness, a product of volume times density," *J. Exptl. Psychol.* 69, 503–510 (1965).
30. G. Ekman and T. Künnapas, "Scales of Masculinity and Femininity. A Further Study of Direct and Indirect Scaling Methods," *Psychol. Lab. Univ. Stockholm Rept. No. 162* (1963).
31. G. Ekman, "Measurement of moral judgment," *Percept. Motor Skills* 15, 3–9 (1962).
32. T. Sellin and M. E. Wolfgang, *The Measurement of Delinquency* (Wiley, New York, 1964).
33. D. Bernoulli, "Exposition of a new theory on the measurement of risk" (originally published in Latin in 1738), *Econometrica* 22, 23–35 (1954).
34. S. S. Stevens, "Measurement, psychophysics and utility," in *Measurement: Definitions and Theories*, C. W. Churchman and P. Ratoosh, Eds. (Wiley, New York, 1959), pp. 18–64; E. Galanter, "The direct measurement of utility and subjective probability," *Am. J. Psychol.* 75, 208–220 (1962).
35. S. S. Stevens, "A scale for the measurement of a psychological magnitude, loudness," *Psychol. Rev.* 43, 405–416 (1936).
36. ———, "On the brightness of lights and the loudness of sounds," *Science* 118, 576 (1953) (abstract).
37. P. L. Pradhan and P. J. Hoffman, "Effect of spacing and range of stimuli on magnitude estimation judgments," *J. Exptl. Psychol.* 66, 533–541 (1963).
38. G. Ekman and L. Sjöberg, "Scaling," *Ann. Rev. Psychol.* 16, 451–474 (1965).
39. The power functions in question were later published in S. S. Stevens, "The psychophysics of sensory function," in *Sensory Communication*, W. A. Rosenblith, Ed. (M.I.T. Press, Cambridge, Mass., 1961), pp. 1–33.
40. J. C. Stevens and M. Guirao, "Individual loudness functions," *J. Acoust. Soc. Am.* 36, 2210–2213 (1964).
41. R. L. Hamblin, D. A. Bridger, R. C. Day, W. L. Yancey, "The Interference-aggression law?," *Sociometry* 26, 190–216 (1963).
42. This article (*Laboratory of Psychophysics Rept. PPR-306*) was prepared with support from NSF grant GB-3211 and NIH grant NB-02974. A slightly modified version will appear in D. K. Whitla, Ed., *Handbook of Measurement in Education, Psychology and Sociology* (Addison-Wesley, Reading, Mass., in press).

Fishbein, M., Attitude Theory and Measurement, 1967, New York.
Hall, E.T., The Hidden Dimension, 1966, New York.
Jackson, D.N., The Dynamics of Structured Personality Tests, 1971, London, Ontario.
Platt, J., What We Must Do, 1969, Ann Arbor