International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

What To Do When The Experts Disagree

Stoto, M.A.

IIASA Working Paper

WP-82-065

July 1982

WHAT TO DO WHEN THE EXPERTS DISAGREE

Michael A. Stoto

July 1982
WP-82-65

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

THE AUTHOR

Michael A. Stoto is Assistant Professor of Public Policy at the J.F. Kennedy School of Government, Harvard University.

ACKNOWLEDGEMENTS

This work was done in conjunction with the IIASA Risk Group's Liquid Energy Gas project, and funded in part by the Bundesministerium für Forschung und Technologie, FRG, contract no. 321/7591/RGB 8001. The author thanks Bruce Ackerman, Brian Arthur, Susan Arthur, Hermann Atz, Jesse Ausubel, David Bell, Howard Kunreuther, John Lathrop, Richard Light, Joanne Linnerooth, Eduard Loeser, Giandomenico Majone, Christoph Mandl, Warren Sanderson, and James Vaupel for discussion and comments. The views expressed are the author's own, and are not necessarily shared by the sponsor or those kind enough to comment.

WHAT TO DO WHEN THE EXPERTS DISAGREE

Michael A. Stoto

INTRODUCTION

Policymakers often call on experts to help assess the facts underlying complex policy decisions. The experts give advice on, for example, the current or future state of the economy, the technological potential of a new research or development effort, the future distribution of population and their demand for services, the ecological effect of a new road, factory, or power plant, or the risk to life and limb associated with a new energy facility. Unfortunately, from the policymakers' point of view, the experts often disagree. Actually, the disagreement is sometimes a blessing in disguise.

The aim of this paper is to consider when and how a policymaker would want experts to agree, how to structure the initial charge to the

experts, and how to deal with the results in order to make the most of their answers. We will use a particular case, three risk assessments for a proposed liquid natural gases (LNG) transfer facility in California, as an example throughout. The problem is of course more general, and we generalize whenever possible. The first section of the paper discusses the issue of whether analytical work and especially formal risk assessment, is useful to policymakers. The next section gives some details about LNG and the California case. We then consider the three main issues: 1) why experts disagree, and when this is desirable, 2) how a policymaker can structure the problem, from the beginning, to get the most out of his expert advisers, and 3) what conclusions a policymaker should draw from a set of disparate risk assessment. The paper concludes with a section on implementing these ideas.

1. THE ROLE OF RISK ASSESSMENT

From the outset, we will need to make a number of assumptions about the role of formal risk assessment in the policy process. The desirability of these proposals has been discussed at length elsewhere, so the arguments behind them will be only briefly sketched.

1. Information helps policymakers. (Hoaglin, et.al., 1982, especially Chapters 1 and 14). The availability of new information can set up conflicts and lead to confusion (Mazur, 1973). Yet in terms of making the best decision for 6society, it is hard to argue that we are better off by sticking our heads in the sand. This position assumes a certain amount of rationality in the policy process.

2. Assessing the risk is only one part of the decision making process. Evaluating the alternatives and choosing among them is the other part (Raiffa and Zeckhauser, 1981). The aim of risk assessment is to supply the policymaker with necessary information to aid the choice, not to suggest a decision. Engineers and scientists are no better than others in deciding *whether* a proposed facility is safe enough; their comparative advantage is telling others *how* safe it really is.

3. It is desirable to separate the assessment of the likelihood of potential consequences from the evaluation of the alternatives. (Raiffa and Zeckhauser, 1981). For example, when deciding whether to build an LNG facility, it is better to explicitly estimate the probabilities of one, ten, or a hundred deaths, and then decide if the economic benefits exceed these human mortality costs, than to combine these two steps. Although this separation may seem difficult and unnatural, it is important because: 1) scientific facts may be complicated, and no single person may have mastered them all, 2) an explicit approach facilitates peer review, 3) analysis is not static, and an explicit approach helps to uncover and fill in weak spots in our knowledge, and 4) the assessment may be valuable to other policy-makers for other decisions.

4. It is helpful to be precise about our uncertainty (Raiffa and Zeckhauser, 1981). Statements like "the risk is low" mean different things to different people, therefore, are really meaningless. It is better for an expert to say "my best guess is ten expected deaths per year, but the real answer could be between one and one hundred with 90 percent probability." An expert who makes a statement like this may be right or wrong, but at least we know what is intended. Winkler (1967), Raiffa

(1968), Savage (1971), Tversky and Kahneman (1974), Hogarth (1975), and Spetzler and Staël von Holstein (1975) discuss methods of eliciting personal probabilities and potential biases in the process.

2. LNG AND THE CALIFORNIA SITING DECISION

As a concrete example of a situation where the experts disagree, let us consider a specific case study by Lathrop and Linnerooth (1982). In 1972, two California utilities created the Western LNG Terminal Company to construct and operate facilities to transfer, store, and distribute liquid natural gas from Alaska and Indonesia.

Natural gas is expensive to transport because of its high volume. One way to make transportation over seas or long distances economically feasible is to condense the gas. Cooling the gas to below -162 degrees centrigrade turns it into a liquid and reduces the volume by a factor of about six hundred (Mandl and Lathrop, 1982) the process involves liquefication plants in the sending areas, special ships and ports to transport and transfer the liquid, and gasification plants in the receiving areas. Western originally proposed to build three facilities to transfer, store and regassify LNG. After an initial round of regulatory and legislative action, only one site, Point Conception, was still under consideration.

One major drawback of LNG use is the potential of a very serious fire causing many deaths. The danger arises when an accident causes a leak in a ship or land based storage tank, or in the transfer facilities. The escaping liquid warms and forms a cloud of flammable gas, which in turn can envelop populated areas and ignite. The number of casualties

depends on how far the cloud travels before it either ignites or dissipates.

The risk to life and limb is obviously one of many factors affecting the decision of whether to build an LNG facility at Point Conception. To assess the magnitude of this risk, and to strengthen their arguments in the regulatory proceedings, various parties commissioned three formal quantitative studies. Western, the company proposing to build the facility, commissioned one study by Science Applications, Inc. (SAI). This study concluded with the statement:

> "As shown, the highest fatality probability is one chance in 14 million per person per year within one and one-third miles of the site, decreasing to probabilities ranging from 1 chance in 1 billion to 1 chance in 10 billion per person per year or less within 2 miles of the site. The probability of one occurence of 10 to 100 fatalities is one chance in 29 billion per year, and the maximum fatality count per occurrence is 54, with a probability of 1 chance in 760 quintillion (760 followed by 18 zeroes) per year." (SAI, 1976, pp. 1-12)

The California Public Utility Commission, required by law to approve LNG sites, commissioned Arthur D. Little, Inc. (ADL) to do another study of the same site. They concluded that:

> "These analyses indicated that the probability of an accident involving ten or more casualties due to the proposed project was around 10^{-8} per year (100 million years recurrence interval) for existing population levels..." (ADL, 1978, p. 13)

Finally, the Federal Energy Regulatory Commission (FERC), which also had to review the plans, conducted its own risk assessment of the Point

Conception and other sites. The tabular summary of their quantitative results include:

Location	Maximum Expected Fatalities per year	Exposed Population	Probability of a Fatality per Exposed Person per year
Point Conception	1.13×10^{-5}	15	7.83×10^{-7}

Here are three expert assessments of the "risk" associated with the same LNG facility. As is most often the case in similar circumstances, the experts do not agree. Let us imagine ourselves as a decisionmaker, say a administrative law judge, who has to decide whether or not to allow Western to build the facility. Among the many factors one must consider is the "risk" of the proposed facility. What should we do in the face of these three reports?

We begin by examining the reasons why the experts disagree. Some of these reasons for disagreement can and should be avoided by prior arrangements with the experts. Others cannot be avoided, and in fact the range of expert opinion sometimes carries useful information. After this, we turn to a discussion of what can be done in advance to make the results of risk assessments most useful. One issue is to what degree the risk assessments should be independent of each other. Another is the appropriate scope of the analysis. We follow this with a discussion of ways to compare and combine the quantitative risk assessments in hand if a decision must be made today. We conclude by suggesting a system to implement these ideas.

3. WHY DO EXPERTS DISAGREE?

When several experts are asked to provide information for policymakers, different experts often provide different information. In some cases, the differences arise from lack of agreement about what the problem is, and proper coordination could reduce the confusion. In other cases, the differences are in some sense "real", and are important for the policymaker to be aware of. A third set of differences reflect bias, and should be avoided. Mazur (1973) discusses a number of these sources of disagreement with regard to expert opinion on the safety of fluoridation of drinking water and low level radiation from nuclear power plants.

There are a number of ways that poor coordination between experts can lead to differences in risk assessments. The three Point Conception analyses provide some concrete examples:

1. *Problem Limits.* While ADL and SAI considered risks associated with vessels, transfer of the LNG, and storage facilities, FERC only considered vessel accidents. Other things equal, this would decrease FERC's estimates of the overall risk. This difference is obvious, but many more subtle versions are common. For instance, analysts must decide whether vessel accidents 5, 50, or 200 miles from the terminal, or even at sea or at foreign ports, should be included. Presumably risks to life are much higher when the ship is near a populated area, but the risk to seamen and other vessels may not be negligible.

2. *Operating Conditions.* SAI made its risk estimate assuming larger ships and more storage tanks than did FERC and ADL. This would naturally increase SAI's risk estimate, but not in a simple way. Other less obvi-

ous differences could be due to timing of ship arrivals, docking procedures, construction standards, and so on.

3. *Background Assumptions.* The number of people assumed to be at risk -- that is living close enough to be affected -- ranged from 15 in the FERC study to 90 in the SAI study. Part of this disagreement was due to different assumptions about future population growth. Assumptions about future traffic patterns at nearby airports and missile ranges are similar sources of differences.

4. *Reporting Language.* Final results can be stated in a number of ways. ADL gives "the probability [per year] of an accident involving ten or more casualities." SAI gives the "fatality probability...per person per year", for people living at varying distances from the site, "the probability of one occurence of 10 to 100 fatalities...per year", and the probability per year of an accident with "the maximum fatality count." FERC's summary statistics are in terms of "maximum expected fatalities per year" and "probability of a fatality per exposed person per year". Often risk statements are in the form of recurrence times rather than probabilities, or probabilities are for periods of one, ten, or twenty years. This problem often cannot be solved by simple mathematical conversion from one language to another; to do that often requires more information than is available to the user of a risk analysis. A more extreme version of this problem is that some experts consider more consequences than others, for instance injuries as well as deaths. Vaupel (1982) points out that any single statistic usually presents a biased picture of the policy relevant information. It is better to choose a small number that fully present the issue.

5. *Conscious Omissions or Conditions.* Often risk assessors consciously decide to omit certain risks, such as sabotage, or condition or certain events, such as the non-occurence of war. To the extent that these omissions or conditions are different from one risk assessment to another, different summary figures will result.

Each of these five types of disagreements is spurious in the sense that it does not reflect real disagreement among the experts about the risk in any given situation. It simply reflects the fact that the experts are considering different situations. Yet the disagreements are disconcerting to policymakers and the public, and can be confused with deeper expert disagreement arising from scientific uncertainty. Furthermore, these differences most often cannot be reconciled retrospectively -- the calculations have to be done almost from scratch if new assumptions are to be used. Proper coordination, that is advance agreement on exactly what situation will be considered and what language will be used, can reduce such disagreements, and focus attention on deeper and more important differences.

A second group of sources of disagreement reflect, in a sense, more basic or underlying scientific uncertainty. Again, the Point Conception case provides some examples:

1. *Models.* Each of the three risk assessments used a basically similar framework to assess the overall risk of an LNG facility. This consisted of studying the conditional probabilities of a chain of events: a ship collision, an LNG spill, vapor cloud formation, dispersion, and ignition, and fatalities. In addition similar models were built to assess the risks associated with transferring and storing the LNG. These models serve as a

framework with which to combine more basic information and data from experience or experiments. But such combining models are necessarily simplifications of reality. And different modelers make different simplifications. Perhaps the most striking example of different models is the description of the dispersion, and ignition of vapor clouds. This is a very complex physical problem which depends on wind, weather conditions, surface geometry, and many other factors. One indicator of the uncertainty associated with choice of model is the fact that, while the ADL, FERC, and SAI models were developed on the basis of the same data on small spills, ADL and FERC find that the downwind dispersion distance increases with decreasing wind speed, SAI finds the opposite. A more subtle example is the assumption of independence between certain events. Presumably the probability of grounding is higher near the shore, and the expected number of people exposed to a vapor cloud is also higher. Yet some models do not consider this interaction (to consider all possible interactions would clearly be impossible) and different risk estimates result.

2. *Sources of Data.* To estimate the probability of an event for which we have little or no direct experience, risk assessors have a range of alternatives. At one extreme they can consider a large data base for events very different from the one under consideration, and make some sort of subjective or analytical adjustment. An example of this would be estimating the probability of a collision based on data for all ships off the California coast. At the other extreme, assessors could use a much more limited number of observations for a more restricted but more similar set of events. For instance, they could only consider tanker accidents in

a more restricted part of the coast. At the first extreme, subjective judgment (as used by ADL and SAI) or explicit modeling (as used by SAI) introduces uncertainty. At the other extreme, uncertainty is primarily due to weak statistical inferences from limited experiences. Thus, both the choice of what data are considered relevant, and the manner of adjustment from limited experience, lead to differences in risk assessment.

3. *Judgmental Probabilities.* For some components of the risk estimation almost no data are available. For instance, FERC and SAI both use models where the probability that the vapor cloud ignites depends on the probability that each person enveloped by the cloud ignites it. But no hard data is available for this latter probability. FERC's judgmental estimate is 0.0025, but SAI's estimate is 0.1, leading to a larger number of fatalities. In general, judgmental estimates are used when data are lacking, so can be a major source of disagreement between risk assessors.

4. *Imagination.* Some experts are simply more imaginative than others, and thus consider a broader range of risk sources. SAI, for example, considered only ship collisions, but not ramming the dock or grounding. ADL considered, but SAI ignored, the probability of storage tank failure to storm and waves. On the other hand, SAI, but not ADL, considered the probability of a simultaneous rupture of more than one tank, due to a common cause, such as an earthquake.

5. *Uncertainty About Consequences.* There is a considerable amount of uncertainty about the physical effects of LNG vapor ignition. There is certainly a thermal effect, and perhaps a blast effect. Both have direct and indirect effects, such as building fires or collapsed buildings. And there is little agreement on how much heat radiation for how long is

necessary to cause a fatality. Since we have no experience with large LNG spills or fires, scientific opinion on these questions is diffuse.

6. *Training and Background.* Scientists from different fields often approach problems in different ways. In particular, they use different models, have access to and use different sources of data, and have different ideas about physical consequences. Some experts may be better than others at developing judgmental probabilities, and some are clearly more imaginative. So because each of these differences leads to variation in risk assessments, drawing experts from a wide range of background will lead to a range of risk assessments.

7. *Statistical Fluctuations.* Estimates of probabilities or consequences based on statistical sampling techniques or on the outcome of experiments exhibit variability due to random sampling fluctuations. Standard statistical techniques such as confidence intervals are useful to describe the magnitude and effect of this source of variability.

Each of the above sources of disagreement between experts represents a sort of true natural-science uncertainty. These are problems of trans-science (Weinberg, 1972) whose resolution is often either impractical or impossible. Decisionmakers need to know the range of this uncertainty for two reasons. First, it may be possible to develop better information for a particularly crucial point of disagreement. An analysis of the reasons for disagreement could help locate such a situation. For instance, the overwhelming source of disagreement among the Point Conception analyses was the probability of a ship collision, which ranged from 1.3×10^{-8} according to SAI to 5×10^{-4} according to FERC. More information on this probability is clearly necessary. But the second, and perhaps

more important, reason to consider natural-science uncertainty applies when there is more than one decision maker. If a range of risk estimates are believable, opposing parties can maintain that one or the other extreme is the best estimate. An understanding of the true range of uncertainty allows us to see whether a party's position is reasonable, or whether it capitalizes on the presence of some uncertainty to justify an unreasonable position.

An example helps here. If we were trying to decide whether to build at a certain site, it would be best to know, with little uncertainty, that the expected number of fatalities is 10^{-6} per year. Then the political debate could weigh this known probability against the benefits of the facility. More realistically, we may find that there is a good deal of natural-science uncertainty, and that the data and models can support any expected number of fatalities between 10^{-8} and 10^{-4} per year. If 10^{-7} is acceptable but 10^{-5} is not, more analysis is necessary. But at least knowing that values higher than 10^{-4} are unreasonable allow us to discount the arguments of a party that implicitly assumes that the risk is 10^{-2}.

The third group of sources of disagreements involve bias of one sort or another:

1. *Intentional Deception.* Since experts like to please their clients, there is an obvious incentive for them to shade the analysis in one direction. And when there is considerable natural-science uncertainty it is easy to do so. Within a reasonable range this is an integral part of the adversary system, but it is possible to go beyond what is reasonable. This type of deception is hard to detect in a complex analysis, and harder to prove. In the Point Conception case, Mandl and Lathrop (1982) have

ranked the three risk assessments on a number of specific issues and found no indication that any one was consistently more conservative. But, as we have mentioned above, the probability of collision is the key variable, and it is estimated by SAI (who represents the utilities) to be less than 10^{-4} as high as estimated by FERC.

2. *Conservative Estimates.* For expedient or benevolent reasons, experts often give conservative risk estimates, that is estimates that they believe are high. For instance, ADL (Summary, page 10) assumes that "all persons within a vapor cloud fire are casualties." Their reasoning is that people inside fire-proof structures might escape injury, but this is hard to model. But if the facility is deemed safe with this conservative assumption, it will surely be safe in actuality. Conservative estimates also arise when assessors do not trust the decisionmakers, so "for the good of society" they say, increase the risk estimates. The problem with conservative estimates is that they tend to cascade (Raiffa and Zeckhauser,1981) and thus yield overall risk estimates much higher than reasonable. An honest decisionmaker will either be misled, and perhaps make the wrong decision, or be forced to compensate somehow for an unknown amount of conservative shading.

Little can be said in favor of either of these biases that lead to differences between risk assessments. Perhaps a good peer review system can help us to detect and eliminate differences due to intentional deception. Openness and a sense of trust, which may be hard to obtain are the solutions to the problem of conservative assessors.

4. SIMULTANEOUS RISK ASSESSMENTS

In risk assessments, as in other policy problems that depend on "expert" opinion, policymakers should request a number of simultaneous, independent, expert studies of the same problem. Individual experts bring preconceptions and biases to any problem. These result from standard methods and assumptions that vary from field to field, and from policy views of the experts or their employer as to how the question should be decided. Simultaneous studies by experts with a wide range of technical backgrounds and policy views will (1) increase the completeness of the analysis by bringing up more of the relevant risks, (2) lead to better estimates of the component risks and thus the overall risk, and (3) give some idea of the certainty or uncertainty of the final results. The first two goals concern better point estimates. The third concerns knowledge of the reliabilities of our results. If every risk assessment in the range of uncertainty would lead to the same policy decision, we could stop. Otherwise, more information could be sought.

One important question is the degree of independence between the simultaneous studies. To be sure, some coordination is desirable. For instance, the experts should agree in advance on the target. In the LEG analysis, this means that they should make the same set of assumptions about the number of ships per year, the physical layout of the facilities, and the population at risk. And the experts should agree in advance on the definitions of risk, be they societal risk, Rasmussen curves, or any combination of the available measures. Preliminary agreement makes the final results more comparable, and helps to focus attention on real differences in expert opinion.

There are two extreme forms of gathering independent expert opinion. In one case, someone could set out a framework for analysis, and groups of experts could fill in estimates of critical quantities. In the LNG risk analysis, this approach would mean, for instance, that the expert would supply the probability of a spill, the probabilities of immediate and delayed ignition, and a probability distribution for distance downwind a vapor cloud can travel. The other extreme is to ask a group independently to formulate the analytical framework,develop the data and estimate the necessary parameters. Depending on the nature of the problem, the optional use of experts will be somewhere between these two extremes.

The first alternative -- independent experts working on a common framework -- has a number of virtues. First, since all of the experts would be dealing with the same variables defined in the same way, peer-review would be straightforward. Second, different people are experts in different aspects of the problem, and such a disaggregated approach could help to focus the work of the experts in their area of expertise. Third, this approach would clearly identify the areas with the most uncertainty, and allow us to target our further research in those areas.

This disaggregated approach could contribute to the first goal of simultaneous studies -- completeness -- by providing a systematic framework within which many experts could search for all of the risks. But it is not easy to deal with a new risk discovered outside of the analytical framework. It helps the second goal -- accurate estimates of the components -- because averages are generally better than single opinions. But the common framework approach does not yield an accurate

estimate of the true range of uncertainty -- the third goal of simultaneous studies. In many problems, and LNG risk analysis is one of them, there is no single appropriate analytical model. Models are approximations of the world, and must by their nature make simplifying assumptions. Different models yield different risk estimates, so relying on the common analytical framework leads to an overly optimistic view of the certainty of the estimate.

One example of this problem is that model builders, for lack of better information or to avoid excess complication, often assume that certain probabilities are independent. In the Point Conception LNG case, it seems that the events of a ship collision and the wind dispersion of a vapor cloud were considered to be independent, and the probability of each was averaged over the distribution of weather conditions. But in foggy weather, the probability of a ship collision is presumably higher, and vapor clouds travel further. Doing the probability calculations with the full distributions leads to higher risk estimates than averaging at each stage. Similar problems arise from assumptions about the type of analytical model, or functional forms for extrapolation.

The main virtue of the second alternative -- independent formulation of the framework and estimation of the parameters -- is that the resulting range of estimates more nearly reflects the true uncertainty in the risk estimate. In addition, the variability of the average of n independent observations is smaller than if the observations were positively correlated. In the extreme, there is no point in paying for more than one study if they all will give the same answer. Of course there are commonly accepted methodologies in the risk assessment business, and common

data sets are often used. But the pressures of an adversarial system, countered by a reasonable peer review of the analyses, will tend to expand the range of estimates towards an accurate assessment of the true uncertainty in the risk estimate. The difficulty with the total independence approach is that different experts structure the problem in different ways, and define different intermediate variables, so only the overall risk estimates are comparable.

The two approaches to the use of experts can be combined in a number of ways. There could be, for example, a small number of independent frameworks proposed, and panels of experts asked to estimate the parameters of each model. Or the work could proceed in stages, with the first step consisting of a panel of experts independently deriving models, and the second stage consisting of each expert estimating the parameters for the other's models. Both approaches would lead to more complete and accurate risk estimates, as well as realistic estimates of the true uncertainty.

In order to coordinate the use of experts, Dalkey and Helmer (1963) have developed the "Delphi Method". This procedure involves the repeated use of an anonymous questionnaire. Feedback on quantitative estimates and their justifications from previous rounds encourages panel members to reconsider extreme estimates. Anonymity prevents strong personality from dominating the group. But the goal of the process is to achieve expert consensus, not to estimate the range of possible views. Press (1978) and Press, Ali and Yang (1979) describe an alternative procedure with only qualitative feedback. This, they feel, relieves pressure for consensus where there is none.

The first goal in risk assessment should be to estimate the risk itself. By facilitating the comparison and aggregation of intermediate and final results, simultaneous risk assessments helps achieve better estimates. The second but not less important goal is to know how reliable the risk estimate is. If the uncertainty is small, the policy decision will depend on other factors, and the arguments will be related to values. If the uncertainty is large, it is possible that more work, say, an experiment, could resolve the differences, and would be worthwhile. More likely, the decision will be made in the face of uncertainty, but all will be better served if the full range of risk possibilities, especially the higher end, is known. Simultaneous risk assessments lead to better point estimates, and some degree of independence leads to a realistic estimate of the uncertainty.

5. THE EAGLE VERSUS THE WORM

Risk assessments are typically done from a "worm's eye" point of view -- the problem is disaggregated into smaller and smaller pieces until the assessment team cannot see the big picture any longer. For instance, the Point Conception risk assessments were based on very detailed models of ship movements, metalurgic studies of storage tanks and their behavior under stress, the dispersion of a vapor cloud over sea and land, weather conditions, and so on. This approach has some obvious benefits, and some might say it is the only way to proceed. But the alternative -- taking an "eagle's eye" view of the problem -- also has some merits, and deserves serious attention.

Perhaps the strongest argument for the worm's eye approach is that there are no overall experts. There may be experts in ship collisions, metalurgy, gas dispersion, and weather conditions, but no individual can be expected to master all of the subtleties of these and other important disciplines. The worm's eye approach allows us to coordinate the expertise of many individual assessors.

Parallel to this argument is the fact that little data exists that is relevant to the big picture. We have very little experience with LNG operations, and less with accidents, but much more that relates to components of the risk assessment process. For instance, by disaggregating we can bring in data on general shipping accidents, chemical plant operations, and the physics of gas dispersion and ignition. The worm's eye approach allows us to make inferences about a complex problem based on a number of specialized data sets.

Disaggregation is also extremely useful in an adversarial situation. The worm's eye approach permits risk assessors to lay out their reasoning in a way that others can follow and thus exposes their analyses to peer criticism. Detailed arguments based on a number of data sets and mathematical models certainly are more conducive to criticism and discussion than a necessarily judgmental eagle's eye risk assessment. Prior coordination of simultaneous risk assessment, as described earlier, helps to facilitate the peer-review process.

Another argument for the worm's eye approach is that the process is not static, and a disaggregated approach lends itself to updating when more or better information becomes available. If a risk assessment consists of a computer model to combine data-based and judgmental

estimates of specific probabilities and consequences, updating simply means re-running the program with a different input value. In the Point Conception case, additional information became available on earthquake risks late in the decision process. With a disaggregated model, this new information should be easy to incorporate.

Finally, the components of a disaggregated model may have other uses. Risk assessments of other LNG facilities will be called for in the future, as well as other operations that involve transporting or storing hazardous material, and some of the components developed for the Point Conception assessment could be easily adopted.

The major problem with the worm's eye approach is that it focuses attention on specific details of the risk assessment and tends to ignore the structure of the combining model. This is particularly dangerous because the way in which the details are combined is potentially, and often actually, more important than the details themselves.

The worm's eye approach, with its reliance on documentation and derivation, favors the use of narrow analytical models and discourages the use of judgment. For instance, model builders favor linear extrapolation of one sort or another to a more speculative judgment-based approach because it is easier to justify and defend, regardless of whether it is more appropriate.

Similarly, the worm's eye view does not tend to give adequate consideration to the way that the parts are combined. All of the Point Conception risk assessments worked with the probability of a ship accident and with the likely dispersion of the resulting vapor cloud. No doubt,

separate experts made good estimates of the relevant probabilities. But the combining models each assumed that these events are independent. This is certainly an easy assumption to make, and makes the calculations and their documentation easy. But because weather conditions and location in the channel effect both probabilities, the assumption of independence is just not right, and could have an important effect on the overall assessment. The problem of *common-mode failures* is a similar example of inadequate combining models common to the worm's eye approach.

Disaggregation leads to a proliferation of numbers and assumptions as well as very complex computer programs. Risk assessment teams only have a fixed amount of time and effort to spend on any project, and having more details must often lead to less careful attention to each individual one. If minor errors in small details tended to cancel, perhaps this inevitable lack of attention would not be serious. But in fact, in complex models, one small slip could have a major effect on the end result. And complex models and computer programs are difficult to verify.

Finally, the worm's eye approach may use experts in a way that is not best suited to elicit their expertise. Dreyfus and Dreyfus (1978) have argued that an expert at his or her best thinks intuitively and holistically, and does not do complex calculations. Only beginners think through every step. Thus asking experts to estimate probabilities and likely consequences for a narrow analytic model is asking them to forsake their expertise.

In risk analysis of low probability events, taking the eagle's eye point of view is equally difficult. Tversky and Kahneman (1974) discuss many subtle biases in assessing probabilities directly. Pratt and Zeckhauser

(1982) discuss biases in probability assessments based on one alarming event. Fairley (1981) has pointed out that even many years of experience with no problems provides little evidence about the size of a small probability. And of course we are talking about facilities that have not yet been built. There is little direct experiential evidence on the risk of LNG facilities.

But there is extensive evidence on the risks of more or less similar industrial activity. Surely, no two industrial experts would make the same judgment about the numerical risk of a proposed facility. But they might be able to give reasonable estimates of the range of uncertainty by making a series of extreme assumptions and comparisons to other types of industrial activity. Mosteller (1977) discusses some helpful procedures for making such order-of-magnitude estimates. The presence of more than one assessment would serve both as a check for extreme assumptions, and as an indication of the degree of certainty. The range of certainty of such an estimate would obviously be large -- five or ten orders of magnitude. But realistic ranges of uncertainty for the more complex models, as we will see later, are equally as large.

When dealing with very small probabilities, the Law of Outrageous Events comes into play. Suppose that a complex risk analysis estimates that the probability of an accident involving ten or more fatalities is about 10^{-11}. Can such a small number be correct? Perhaps it can be, conditional on the assumptions of the model. But then there always is the possibility that an outrageous event will occur. In the LNG case, sabotage is an obvious excluded possibility, but so is a tidal wave. Each of these may have small probabilities, but compared to 10^{-11}, they may be large.

The point is that when the worm's eye approach comes up with a very low probability, we must begin to explore possibilities that would otherwise not be worth considering.

There is, of course, no reason to take only one approach. A detailed worm's eye analysis provides many insights about the risk problem, and may lead an assessor to a better understanding of the situation. But the estimates should be tempered with an independent eagle's eye consideration of whether the magnitude of the results are reasonable.

6. COMPARING AND COMBINING RISK ANALYSES

The previous discussion concerns steps to be taken in advance of or during the work of risk assessment teams to make the results more comparable and suitable for public decision making. But often decisions must be made on short notice with no chance to gather new data or modify old analyses. Faced with a decision and a small number of inconsistent risk analyses, what should a decisionmaker do?

Our approach here assumes that experts are unbiased, and try their best to provide probability assessments that are independent of their policy views. But even so, there will be differences due to natural-science uncertainties and perhaps lack of coordination. So we will behave as if we have multiple estimates of the same quantity, as if chosen as samples from the same (subjective) probability distribution. The goal is to assess and summarize the underlying distribution of expert opinion. The supposition that experts can give estimates that are not influenced by policy views may be optimistic, but it at least provides a starting point, and

perhaps a goal to strive for. In the next section we consider modifications when we are not confident that the expert assessments are independent of their policy views, and tend to group in two schools.

1. *Pick a Favorite.* Dealing with conflicting information is difficult, so policymakers are often tempted to simply pick a favorite study and ignore the others. If there were only one appropriate approach or answer, this strategy could possibly work, but as we have seen, this is not the case. The trans-science aspect of risk analyses and most other public policy problems implies there is no single appropriate answer, but a range of possible answers. Reporting a single number tends to hide the important fact that there is substantial scientific disagreement. Even if there were a single best answer, it is not clear how a policy maker could find it among the pack.

2. *Average the Results.* Another approach is to average the final results. For instance, the "societal risk" estimates for Point conception were as follows (Mandl and Lathrop, 1982):

SAI	$1 \cdot 10^{-6}$ expected fatalities per year
ADL	$7 \cdot 10^{-6}$ expected fatalities per year
FERC	$1 \cdot 10^{-5}$ expected fatalities per year

The average of these numbers is 6.0×10^{-6} expected fatalities per year. But this procedure makes a number of questionable assumptions. First, it gives each assessment equal weight. It is quite conceivable that we would want to give more weight to the more reliable experts, in the way that we use weights that are inversely proportional to variance to yield the most efficient statistical summaries. But it is not clear how to derive these

weights, short of extensive precious experience with the experts. And assessing an expert's track record is exceptionally difficult if the goal has been to estimate very small probabilities. DeGroot (1974) gives one approach to developing such weights based on the expert's opinions of one another. Hogarth (1975) reviews other methods, and concludes that equal weights often perform well compared to self-ratings or past performance. Second, averaging does not take into account any information we may have about potential sources of bias. In addition, there is a question of scale. For small risks, the order of magnitude is the crucial issue, therefore, the logarithmic scale is appropriate (Hofstadter, 1982). For the three Point Conception, estimates averaging in the logarithmic scale, or, equivalently, using a geometric mean, yields 4.1×10^{-6}, which differs slightly from the straight average of 6.0×10^{-6}. If the risks are substantially different in order of magnitude the choice of scale makes a difference. For instance, the straight average of 10^{-8} and 10^{-4} is 5.0005×10^{-5}, fifty times higher than the geometric mean of 10^{-6}. If order of magnitude is the key question, the second approach seems to yield a more natural summary. The main benefit of averaging is protection from reliance on a single estimate that could turn out to be unrealistic. An alternative is to use the median assessment, which is not overly affected by one outlying estimate.

3. *Bayesian Updating*. Morris (1974, 1977) and others have suggested Bayesian updating as a way of combining expert evidence. The basic idea is that each expert assessor i would make an estimate of the risk, p_i, and that the decisionmaker would combine these with an *a priori* subjective distribution on the true risk, π. This approach has a number of

problems. First, the decisionmaker must make explicit and use *a priori* distributions for π. Policymakers are not used to thinking in these terms, and there are obvious problems with multiple parties with different prior ideas. Second, to use Bayes' Law, the decisionmaker must have a subjective distribution for what each assessor will say, given the true risk π. This is one way to build in some prior notion of potential bias, but is extremely difficult to quantify. Given this complexity, Bayesian updating does not seem to offer a practical solution to the problem of disagreement between experts.

4. *Sample Based Distribution Assessments.* If the separate assessments can be regarded as independent estimates, the range of estimates can be used as a guide to the true uncertainty of expert opinion. An example from the Point Conception case helps here. Both ADL and SAI, but not FERC, made calculations of the annual probability of ten or more fatalities due to an LNG accident. The high estimate was ADL's at 1.0×10^{-8}, and the low estimate was SAI's at 2.2×10^{-11}. A policymaker's initial thought may be that this interval must nearly cover the range of possible expert opinion. In fact, there may be a substantial probability that a new expert would make an estimate outside the interval.

Because of the scale problem, let us work with logarithms to base 10. In this scale, ADL's estimate is -8.00 and SAI's is -10.66. Let us assume that there is some true risk, but that, because of uncertainty in the estimation process, the assessments can be regarded as independent observations from a Normal distribution centered around the true risk. The object is to estimate the center of the distribution, the true risk, and the variance, a measure of the uncertainty. With these assumptions, the best

estimate of the true risk is the average of the two observations, −9.33, or converted back into probabilities, $4.7x\,10^{-10}$. For a sample of size two, the standard deviation can be estimated as 0.886 times the range, that is, 2.30. In terms of percentiles, the subjective distribution based on these calculations, in terms of probability, is:

Percentile	10	25	50	75	90
Probability of Ten or More Fatalities	$4.5x10^{-13}$	$1.2x10^{-11}$	$4.7x10^{-10}$	$1.8x10^{-8}$	$4.9x10^{-7}$

This range is quite large, so one would say that there is not much certainty about the risk of the proposed plant. The major assumption here is that the two risk assessments are independent. If the estimates were made for or by opposing parties, there might be a tendency for one to be too high and the other too low. In this case, the calculated range would be too large. Similarly, if the two assessors were biased in the same direction, the calculated range understates the true range of uncertainty.

If there are a number of simultaneous, independent, disaggregate assessments, the same approach can be extended. For instance, the annual probability of more than ten fatalities (due to a ship accident) can be estimated as the product of a series of conditional probabilities: the probability of a ship collision, the probability of a spill given a collision, the probability that a vapor cloud forms given a spill, and the probability of a blast or fire killing more than ten people given the formation of a cloud. According to the ADL and SAI risk assessments, the probability of this chain of events in much higher than other chains leading to similar accidents. There are, of course, other ways to specify the chain, but this

particular description allows easy comparison of the three Point Conception risk assessments. Table 1 gives the estimates of these probabilities taken from each of the reports. FERC does not calculate the last conditional probability.

Table 1

-	ADL	FERC	SAI	Geometric Mean
P(Collision)	9.5×10^{-4}	8.8×10^{-2}	7.3×10^{-6}	8.5×10^{-4}
P(Spill|Collision)	8.0×10^{-2}	4.5×10^{-3}	2.5×10^{-1}	4.5×10^{-2}
[P(Spill)]	7.6×10^{-5}	$[4.0 \times 10^{-4}][1.8 \times 10^{-6}]$	$[3.8 \times 10^{-5}]$	
P(Cloud|Spill)	0.1	0.1	0.1	0.1
P(>10 death|spill)	1.3×10^{-3}	--	1.2×10^{-4}	3.9×10^{-4}
P(>10 deaths)	1.0×10^{-8}	--	2.2×10^{-11}	1.5×10^{-9}

Sources: ADL 1978a, pp. 5-4, 5-21; ADL 1978b, p. 13; FERC 1978, p. 533; SAI 1976, pp. 1-6, 1-12, 5-31, 8-149; and calculations.

First, by averaging each conditional probability, we can take account of the FERC estimates of the first two parts of the chain. The product of the mean conditional probabilities is 1.5×10^{-9}, slightly more than the 4.7×10^{-10} average of the ADL and SAI final estimates. Second, if the individual conditional estimates are independent, and we continue with the assumption that the uncertainty in all of the estimates has a Normal distribution in the log scale, we can calculate the uncertainty of each component of the estimate, and calculate the joint effect on the final result.

First note that estimates of the probability of a collision and the conditional probability of a spill seem to be negatively correlated: FERC gives

the lowest probability of collision but the highest conditional probability of a spill, and SAI is just the opposite. Most likely, FERC was generous in what it labeled a "collision", so had a higher probability of collision, but a lower probability of a spill. For this reson, let us multiply the two together, and use instead the unconditional annual probability of a spill.

Let P be the probability of ten or more deaths. This is the product of three factors: Q, the probability of a spill; R, the conditional probability of a cloud forming; and S, the conditional probability of ten or more deaths. Each is assessed with some uncertainty; let σ_Q^2 , σ_R^2 , and σ_S^2 be the variance of the logarithms of Q , R , and S respectively. Then if the assessments of Q , R , and S are independent, σ_p^2, the variance of the logarithm of P is $\sigma_Q^2 + \sigma_R^2 + \sigma_S^2$. All three estimates of R are the same, so we might say that $\sigma_R^2 = 0$. More realistically, there should be some uncertainty, but it will be small compared to the other two components. The common logarithms of the maximum and minimum estimate of Q are -3.40 and -5.74. For three independent observations, .591 times the range is an estimate of the standard deviation. Thus $\sigma_Q = .591(2.34) = 1.39$. Similarly, for two independent observations, $\sigma_S = .886(1.03) = 0.91$. Thus, the variance of P is

$$\sigma_p^2 = \sigma_Q^2 + \sigma_R^2 + \sigma_S^2 = (1.39)^2 + 0^2 + (0.91)^2 = 2.76$$

and $\sigma_p = 1.66$.

Based on these calculations, the quantiles of the subjective distribution for the probability of an accident are:

Percentile	10	25	50	75	90
Probability of Ten or More Fatalities	1.1×10^{-11}	1.1×10^{-10}	1.5×10^{-9}	2.0×10^{-8}	2.0×10^{-7}

The simplicity of this approach comes from the assumptions of the logarithmic scale for for multiplicative probabilities and independence of assessors and components of athe assessment. More complex probability models or computer simulations could be used for the same purpose if these assumptions did not hold.

Subjective Distribution Assessments. A final way to deal with a set of divergent expert opinions is to hire another expert to review the reports and to report a subjective probability distribution summarizing the risk estimates. The new expert could use some of the techniques mentioned above, but could also combine the information in a less formal way, and take other factors into account. The role of the new expert is to provide a "best" estimate of the risk, and more important, to define a range of "reasonable" values to focus the insuing political decision on values rather than facts. It is obviously important to find an unbiased expert to combine the various opinions, but since we seek a range of reasonable values, rather than a single best number, such a process is possible. Arthur (1982), for instance, provides such a review of current estimates of world oil resources.

As an example of this process, I asked two IIASA colleagues to provide their subjective probability distribution on the annual probability of an accident involving ten or more fatalities. Each expert has a technical background and has worked closely with the three Point Conception risk

analysis reports, as well as the others studied in the IIASA LNG risk project. To assess their subjective distributions, I used the methods described by Spetzler and Staël von Holstein (1975) and Morgan, Henrion and Morris (1979). These include an initial discussion with the experts concerning their knowledge of the situation, their biases, exactly what probability is being estimated, implicit conditions, and in what scale they feel most comfortable working. For instance, the first expert felt comfortable directly assessing the annual probability of an accident involving ten fatalities due to any cause, including sabotage. The second was more comfortable separately assessing the probability of such an accident under normal conditions and due to sabotage, and wanted to give probabilities for a fifteen year period. By simple probabilistic calculations, I was able to convert the second expert's distribution into terms consistent with the first's. The 10th, 25th, 50th, 75th, and 90th percentiles are as follows:

Percentile	10	25	50	75	90
Expert 1	$1{\cdot}0\times10^{-10}$	$5{\cdot}0\times10^{-7}$	$1{\cdot}3\times10^{-5}$	$2{\cdot}0\times10^{-4}$	$1{\cdot}0\times10^{-3}$
Expert 2	$4{\cdot}0\times10^{-7}$	$5{\cdot}0\times10^{-6}$	$3{\cdot}2\times10^{-4}$	$2{\cdot}0\times10^{-3}$	$7{\cdot}8\times10^{-3}$

The fact that these distributions are reasonably close -- they differ by about one order of magnitude -- suggests that the process of assessing subjective distributions is reliable and gives an honest assessment of the best estimates of the risk (about 10^{-4}) and the range of reasonable disagreement among experts (about 10^{-9} to 10^{-2}). These results are substantially different from and higher than the estimates based on mechanical combination of the individual estimates. Because the subjective results are both more complete (they include sabotage, for instance) and

more considered, they are probably more realistic.

The techniques discussed here relate to estimating and reporting the distribution of expert opinion for single quantities. Decisionmakers of course need more information. First, as we have discussed above, no single number carries all of the policy relevant information. Policymakers want to know about the expected number of fatalities, probabilities of small accidents and major disasters, separate estimate of the probability of sabotage, and so on. Different parties are concerned about different aspects of the problem. Thus, the distribution of expert opinion should be simultaneously assessed for a number of quantities. Second, decisionmakers need to know the reasons for expert disagreement as well as the range of possible values. Knowledge of the reasons for disagreement helps us estimate the likelihood that more effort (an experiment, for example) would produce agreement, and also helps us to choose among the estimates if a choice must be made.

There is of course no single correct technique for combining and comparing expert opinion. The methods discussed here all have their strengths and weaknesses. Perhaps it is best to try as many approaches as possible, and then attempt to understand why they differ. In any case, whoever does the combination and comparison should remember that assessing the range of disagreement is as important as getting the single best estimate.

7. BIASED EXPERTS

The techniques in the previous section rely in a number of places on one crucial assumption, that experts are unbiased. In statistical terms, we can be more precise, and say that the process of obtaining an expert assessment is like drawing a sample from a distribution. We have assumed up until now, that the expert opinions were all being drawn from the same distribution. But there are a number of reasons to question that assumption, and explore consequent changes in the procedures for combining expert opinion.

There are at least two situations that could lead to bimodal distributions of expert opinions. One involves a natural-science uncertainty, and the other arises when experts can not separate assessments and values. Of course, the two may be related.

In some situations, one single natural-science uncertainty overrides all of the others. For instance, one of the key factors in assessing the risk associated with the disposal of nuclear wastes the biological effects of low level radiation. Most of our information about the effects of radiation comes from studies of animals or humans exposed to relatively high doses of radiation. Dose-response curves then provide a means of extrapolation to lower doses. But there is little agreement on the appropriate shape of such curves. The choice of a linear model over a threshold model in a risk estimate implies a difference of many orders of magnitude. The effect of other modelling choices is small compared to this one factor. Thus if one school of experts believes in the linear model, and a second school believes in some other model, risk estimates will tend to cluster in two groups. It would be misleading to summarize the expert

opinions with a single number, or two assume a simple unimodal distribution.

As much as we would like to think that experts can divorce themselves from values in making their evaluations of the probabilities and consequences of potential decisions, this is often not possible. Scientists are also men and women of the world, and like everyone else, have views on policy matters. Even if they try to make estimates to the best of their ability it is likely that subtle biases will creep in whenever judgement is called for. Experts of like persuasion tend to associate with one another, and thus be exposed to similar ideas about appropriate models or data. The net result of shading a number of factors in the same direction is a bimodal distribution of the final estimates. And if there is a single dominant natural science uncertainty as discussed above, experts with similar political views will tend to group in one extreme or the other. Of course, if experts are acting as advocates, these tendencies will be even stronger.

If expert opinions are bimodally distributed, picking a favorite implies choosing one school and totally ignoring the other. Similarly, averaging the estimates is also misleading. First, a single average hides the important fact that there really are two divergent points of view. A number in the center in not regarded as correct by either school. Second, if the experts really form two distinct groups, averaging is like voting: the relative number of opinions in each group is the crucial factor. But there is usually no reason to suspect that the number of experts in each group, either in the sample, or in the population, has any meaning. The fact that three times as many experts take one position as

another does not mean that the first position is more likely to be correct. Of course this reasoning does not go on forever -- if only one out of one hundred scientists believe in a position, we do have reason to be suspect. Perhaps a better alternative to a single average is one average for each group, if they can be identified as such, and the number of experts taking each polar position.

As discussed in the previous section, Bayesian updating theoretically offers a means of correcting for bias, but there are a number of difficulties. First, the decisions maker, if there is a single one, must be able to specify probabilistically the extent of each expert's bias. This is obviously a difficult task for someone unskilled in the language of probability. But more importantly, decision makers have no way of knowing the magnitude of an expert's bias, even if they can guess the direction. And if experts knew that they were being second-guessed, they might try to overreact, and thus hopelessly confuse the situation, or simply refuse to participate.

The techniques for assessing the distribution of expert opinion that were discussed earlier are strongly dependent on the assumption of independence, so are not appropriate for experts who are biased. One alternative is to assume that the range of expert opinions corresponds to the range of possible values. The two groups would have the tendency to move as far apart as possible. But the decisionmakers who have to use the information have no way of knowing how extreme the expert positions are. They do not know, for instance, whether the probability that the risk exceeds the highest value presented is 0.1 or 0.001. It all depends on the zeal of the experts.

Perhaps a subjective distribution assessment is the best alternative, but even that has a number of difficulties. Perhaps a consideration of the political stances of the experts together with detailed review of their reports could produce an informed view of the extent of bias in their assessments, and suggest a realistic range of possibilities. At the least it could help the decisionmakers to realize that there are two polar positions on the matter, and help to sort out the likelihood of each. But if there are no unbiased experts to provide the inputs, it will be difficult to find an unbiased, but informed, expert to combine the results.

Bias is a serious problem in risk assessments primarily because we do not know its magnitude, even though we may suspect its direction. We thus do not know what we are buying. Finding experts who can report estimates that are independent of policy views, and urging them to try, tends to make their results easier to interpret, and thus much more useful.

8. COORDINATION AND REVIEW

Up until now, the discussion of how to structure the interaction with experts and combine their results has been quite optimistic. It has assumed that information on the probabilities and consequences of accidents would help the decision-makers, and that it is a good idea to separate the functions of assessing and evaluating risks. But matters are not that simple.

For one thing, where there are multiple risk assessments, they are often not done simultaneously. Instead they are commissioned by the

various interested parties as they become necessary. This has two important ramifications. First, since plans for the facility naturally evolve over time, operating assumptions, and the resulting risk assessments, will differ. Second, if one study is already in the public domain, it is hard for subsequent analysts to be independent.

Perhaps a more basic difficulty is that the assessments are often commissioned or even carried out directly by the parties themselves. Even if one agrees in principle that the functions of assessment and evaluation should be separate, it is hard to resist the temptation to shade questionable judgments at every stage. But as we have seen, opposing biases does not necessarily lead to good summary measures, or informative estimate of the true uncertainty. These problems are of course worse if the parties to the decision are unsymetrically supplied with experts.

In addition, once the assessments have been performed (no matter how), someone must compare, combine, and translate them for the decisionmaker. Risk assessments are very complex, and their reports are often exceptionally difficult for even a trained scientist to read. Summarizing this information requires a substantial amount of judgment, so the question of bias again appears.

One solution of these problems is to have some sort of impartial board or arbitrator to coordinate the experts before they do their assessments, and to compare, combine, and translate the results into plain English afterwards. The responsibility of this board would be to lay out what is known in an impartial and informative manner for all of the interested parties, and to define the range of "reasonable" assessments.

Policy decisions based on this common information would then reflect differences in how the parties value the alternatives, not on differences of opinion of the probabilities and consequences involved.

One might ask how unbiased individuals could be found to perform this function, and there are a number of possibilities. In labor negotiations professional arbitrators are often called in to settle disputes. Their job is harder since they have to deal with values as well as facts, so similarly unbiased technical risk assessors should be available. Perhaps academics, preferably from another part of the country, would be a good source of unbiased information coordinators.

Ackerman *et al.* (1974) have proposed an independent board composed of technically trained individuals to review analytic studies for policy decisions. This board's objective is to assess the analyses in plain English on four dimensions: 1) the empirical basis for the report's findings, 2) the extent that the technical discussion diverts attention from other factors, 3) the "scientific competence" of the analysis, and 4) the inherent limitations of the approach. To this we would add the functions of choosing the experts and coordinating their work. Ackerman *et al.* suggest that if the jurisdiction of the board is wide, it will be difficult for a single interest group to capture the board, or pack it with sympathetic members -- as the number of issues increases, it becomes harder to find analysts whose technical learnings all correspond with an interest group's policy views. The "product" of the review board would be a published report aimed at the decisionmaker, but available to all interested parties. If all agreed in advance that the review board's report would define the territory for the subsequent policy battle, all sides (including poorly

financed opposition groups) would have access to informative, reliable and usable risk assessments.

In the California case (Lathrop and Linnerooth, 1982), an administrative law judge was the final arbiter between the applicant, the federal regulatory agencies, and the local residents. This judge or his staff would have been the logical one to convene such a board. With the increasingly sophisticated scientific and technological arguments in regulatory cases today, it would not be unreasonable to build permanent staff expertise for coordinating expert evaluations.

The three other cases studied by the IIASA group offer similar potential locations for the coordinating function. In the Netherlands (Schwartz, 1982) the question of whether to build an LNG facility in Eemshaven was eventually decided by the national cabinet because of the large number of issues involved. The decision to build LNG and associated facilities in Mossmorran in Scotland (Macgill, 1982) was eventually decided by the UK Secretary of State for Scotland. In fact, both the Dutch cabinet and the Scottish Secretary of State did commission single expert risk assessments that were used by both sides. The studies could have been improved by asking for a small number of independent, simultaneous, quantitative studies.

In the Federal Republic of Germany (Atz, 1982) a number of narrow risk assessments were made by various independent experts at early stages of the debate about an LNG facility in Wilhelmshaven. But before the Federal Ministry of Transportation took its final decision, all of the expert studies were reviewed and analyzed by a working group of the Advisory Committee for the Transportation of Hazardous Goods. This

committee is a permanent board of experts for the Ministry. Even though four of the five members of the working group had already been involved in the decision process, the committee was able to reach a concensus.

The basic point is that in most cases where expert opinion can help inform policymakers even though many parties have a say, there is still a single individual or agency changed with the final decision. All of the parties would be well served if this single "pointman" coordinated the information gathering, thus focused attention on the political evaluation of the alternatives, not their technical assessment.

CONCLUSIONS

Making good use of experts in a policy decision requires planning and coordination. If we agree that the role of experts is to inform, not to decide, then policymakers must take a number of steps before and after the experts do their work.

Before they begin their work, someone must coordinate the experts so that they are working on the same problem. There are enough real sources of disagreement and no reasons to add spurious ones. Second,the experts should work independently. This leads to better estimates of the true risk, and just as importantly, to a realistic idea of the range of uncertainty. Technical, model based assessments and subjective judgments both have a role. Finally, it is important to use experts who can report honestly on their assessment of the scientific facts and uncertainties. Although bias is hard to avoid, it leads to confusion in interpreting the expert's assessments, and should be reduced wherever possible.

After the experts have communicated their results, hard work is still required to distill their varied conclusions into a single report. Simple methods like averaging help to obtain a single number, but ignore the range of uncertainty. Because we may want to obtain new information, or set bounds on reasonable arguments, it is just as important to report the uncertainty as the best estimate. This is especially true if there are two or more discrete schools of experts. Mechanical and subjective combinations of the individual results can convey to policymakers an accurate picture of what and how much the experts really know.

Most policy decisions, although they involve many parties, are ultimately decided by one person or committee. This focus could provide a good location for a technically trained "expert coordinator." This person or committee could serve to both coordinate the work of experts in advance, and to combine and compare their conclusions in the end. The effect would be a better informed policy process, and one in which the arguments concerned the values that parties place on the various proposals, and didn't exploit scientific uncertainly for political purposes.

REFERENCES

Ackerman, Bruce A., Susan Rose Ackerman, James W. Sawyer, Jr., and Dale W. Henderson (1974). *The Uncertain Search for Environmental Quality*. New York: The Free Press.

Arthur D. Little, Inc. (1978a) LNG Safety Study, *Technical Report No. 16 in Support of Point conception Draft Environmental Impact Report*, C-80838-50, Cambridge, Mass.

Arthur, D. Little, Inc. (1978b). *Draft Environmental Impact Report for Proposed Point Conception LNG Project*, C-80838-50, Cambridge, Mass.

Arthur, Susan (1982). "Oil Resource Estimates: How Much Do We Know". Working Paper, WP-82-20, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Atz, Hermann (1982). "The Federal Republic of Germany Case Study", in:

Risk Analysis and Decision Processes. Howard Kunreuther and Joanne Linnerooth (eds.), forthcoming from Springer-Verlag.

Dalkey, Norman C. and Olaf Helmer (1963). "An Experimental Application of the Delphi Method to the Use of Experts", *Management Science*, 9, pp. 458-467.

DeGroot, Morris H. (1974) "Reaching a Consensus", *Journal of the American Statistical Association*
69, pp. 118-121.

Dreyfus, Hubert L. and Stuart E. Dreyfus (1978). "Inadequacies in the Decision Analysis Model of Rationality", in: C.A. Hooker, J.J. Leach, and E.F. McClennen (eds.), *Foundations and Applications of Decision Theory*, Vol.I, Dordrecht, Holland: D. Reidel, pp. 115-124.

Fairley, William B. (1981) "Assessment for Catastrophic Risks", *Risk Analysis*, 1, pp. 197-204.

Federal Energy Regulatory Commission (1978) *Western LNG Project Final Environmental Impact Statement*, FERC/EIS-002F, Washington, DC.

Hoaglin, David C., Richard J. Light, Bucknam McPeek, Frederick Mosteller, and Michael A. Stoto (1982) *Data for Decisions: Information Strategies Policy Makers*, Cambridge, Mass.: Abt Books

Hofstadter, Douglas R. (1982). "Metamagical Themas", *Scientific American*, 246, May, pp. 16-23.

Hogarth, Robin M. (1975). "Cognitive Processes and the Assessment of Subjective Probability Distributions", with discussion, *Journal of the American Statistical Association*, 70, pp. 271-294.

Lathrop, John and Joanne Linnerooth (1982). "The United States Case

Study", in: *Risk Analysis and Decision Processes*, Howard Kunreuther and Joanne Linnerooth (eds.), forthcoming from Springer-Verlag.

Mazur, Allan (1973). "Disputes Between Experts", *Minerva*, 11, pp. 243-262.

Macgill, Sally (1982). "The United Kingdom Case Study", in: *Risk Analysis and Decision Processes*, Howard Kunreuther and Joanne Linnerooth (eds.), forthcoming from Springer-Verlag.

Mandl, Christoph and John Lathrop (1982). "Assessment and Comparison of Liquefied Energy Gas Terminal Risks", in: *Risk Analysis and Decision Processes*, Howard Kunreuther and Joanne Linnerooth (eds.), forthcoming from Springer-Verlag.

Morgan, M. Granger, Max Henrion, Samuel C.Morris (1979). *Expert Judgments for Policy Analysis*, Brookhaven National Laboratory, BNL 51358, UC-13.

Morris, Peter A. (1974) "Decision Analysis Expert Use", *Management Science*, 20, pp. 1233-1241.

Morris, Peter A. (1977). "Combining Expert Judgments: A Bayesian Approach", *Management Science*, 23, pp. 679-693.

Mosteller, Frederick (1977). "Assessing Unknown Numbers: Order of Magnitude Estimation", in: *Statistics and Public Policy*, William Fairley and Frederick Mosteller (eds.), Reading Mass.: Addison-Wesley.

Pratt, John W. and Richard Zeckhauser (1982). "Inferences from Alarming Events", *Journal of Policy Analysis and Management*, 1, pp. 371-385.

Press, S. James (1978). "Qualitative Controlled Feedback for Forming

Group Judgments and Making Decisions", *Journal of the American Statistical Association*, 73, pp. 526-535.

Press, S. James, M.W. Ali and Chung-Fang Elizabeth Yang (1979). "An Empirical Study of a New Method for Forming Group Judgments: Qualitative Controlled Feedback", *Technological Forecasting and Social Change*, 15, pp. 171-189.

Raiffa, Howard (1968). *Decision Analysis*, Reading, Mass.: Addison-Wesley.

Raiffa, Howard and Richard Zeckhauser (1981). "Reporting of Uncertainties in Risk Analysis", draft.

Savage, Leonard J. (1971). "Elicitation of Personal Probabilities and Expectations", *Journal of the American Statistical Association*, 66, pp. 783-801.

Schwarz, Michiel (1982). "The Netherlands Case Study", in: *Risk Analysis and Decision Processes*, Howard Kunreuther and Joanne Linnerooth (eds.), forthcoming from Springer-Verlag.

Science Applications, Inc. (1976). *LNG Terminal Risk Assessment Study for Point Conception, California*, SAI-75-616-LJ, La Jolla, California.

Spetzler, Carl S. and Carl-Axel S. Staël von Holstein (1975). "Probability Encoding in Decision Analysis", *Management Science*, 22, pp. 340-358.

Tversky, Amos and Daniel Kahneman (1974). "Judgment Under Uncertainty: Heuristics and Biases", *Science*, 185, pp. 1124-31.

Vaupel, James W. (1982). "Statistical Insinuation", *Journal of Policy Analysis and Management*, 1, pp. 261-263.

Weinberg, Alvin M. (1972). "Science and Trans-Science", *Minerva*, 10, pp. 209-222

Winkler, Robert L. (1967). "The Quantification of Judgment: Some Methodological Suggestions", *Journal of the American Statistical Association*, 62, pp. 1105-1120.