Acceptable Risk:
A Conceptual Proposal*

Baruch Fischhoff**

Introduction: The Search for Acceptability

Perhaps the most widely sought quantity in the management of hazardous technologies is the acceptable level of risk.1 Technologies whose risks fall below that level could go about their business, without worrying further about the risks that they impose on others. Riskier technologies would face closure if they could not be brought into compliance. For designers and operators, having a well-defined acceptable level of risk would provide a clear target for managing their technology. For regulators, identifying an acceptable level of risk would mean resolving value issues at the time that standards are set, allowing an agency's technical staff to monitor compliance mechanically, without having to make case-specific political and ethical decisions. For the public, a clearly enunciated acceptable level of risk would provide a concise focus for evaluating how well its welfare is being protected -- saving it from having to understand the details of the technical processes creating those risks.

A recent example is the attempt by the Environmental Protection Agency (EPA)2 to cope with a case involving vinyl chloride.3 EPA interpreted that case to require it, first, to assess the health risks for emissions of a particular pollutant and, second, to determine an acceptable risk for a source category emitting that pollutant.4 An earlier example is the Nuclear Regulatory Commission's5 attempt to set the overall level of safety expected for nuclear power plants.6 Both are still in process, and it is premature to judge their outcomes.

Yet, an ominous sign may be found in an EPA study done to prepare for dealing with the vinyl chloride case. It "surveyed a range of health risks that our society faces" and reviewed acceptable-risk standards of government and independent institutions.7 This led EPA to find that "No fixed level of risk could be identified as accept-able in all cases and under all regulatory programs...," and that:8

...the acceptability of risk is a relative concept and involves consideration of different factors. Considerations in these judgments may include: The certainty and severity of the risk; the reversibility of the health effect; the knowledge or familiarity of the risk; whether the risk is voluntarily accepted or involuntarily imposed; whether individuals are compensated for their exposure to the risk; the advantages of the activity; and the risks and advantages for any alternatives.

To regulate a technology in a logically defensible way, one must consider all its consequences, i.e., both risks and benefits. To regulate in an ethically defensible way, one must consider its impact on individuals, as well as on society as a whole.

An analytical procedure is advanced here to meet these constraints in determining the acceptability of technologies, one that is consistent with court decisions and compatible with general public values. The next section formulates this concept more precisely. It is followed by a discussion of how it could be implemented procedurally and describes modest compromises to the absolute principle to make it practicable. Embedded in an acceptable political process, the suggested procedure would offer some chance of making the regulation of hazardous technologies more predictable and satisfying. Along these lines, the final part of the paper speculates on how this procedure would affect the fate of particular technologies.

In an essay such as this, it is impossible to work out all the details; the proposal should be judged by whether the concept makes sense and whether its implementation seems workable. It should be appraised in absolute terms: How well could it ever work? What degree of closure would it provide? It should also be considered relatively (recognizing the opportunities competing approaches have had to be proven or discredited): How does it compare to what we have?

This proposal tries to implement the non-utilitarian principle that a technology must provide acceptable consequences for everyone affected by it. Pursuing it as far as possible should produce a better regulatory process than current approaches -- ones focused on other ethical principles (or no explicit principles at all).

If the proposal is attractive, then one might undertake the chore of working out its details. That would involve some daunting challenges, e.g., estimating the scientific uncertainty regarding the magnitude of a technology's risks and eliciting citizens' willingness to trade off diverse costs and benefits.

It will be argued here that such obstacles are a sign of strength rather than weakness. They are inherent in analytically defining acceptable risk and revealed most clearly by an approach that attempts to address them head on. Short cuts can have both direct costs (e.g., antagonizing those whose issues are ignored) and opportunity costs (e.g., keeping scientists from working on neglected issues).

A final proviso is that the proposal would provide an incomplete path to regulatory reform even if all its methodological problems were solved. An analytical principle for evaluating the acceptability of technologies may become a point of departure for struggles, possibly involving suits, lobbying, hearings, demonstrations and negotiations. An analytical approach to acceptability can only hope to forestall some conflicts, by identifying politically unacceptable solutions, and focus others, by concentrating attention on critical unresolved issues. However attractive its logic, an analytical approach will aggravate controversy if offered as a substitute for an acceptable political process. People quite legitimately care as much about how decisions are made as what decisions are made.9

A Proposal for Acceptability: Balancing Risks and Benefits

As EPA has noted, the acceptability of a risk depends on many factors. In their everyday lives, people do not accept or reject risks in isolation. Rather, they make choices among courses of actions, whose consequences may include risks. If people accept a course of action, like deciding to drive somewhere, despite knowing about risks, then those risks might be termed acceptable in the context of the other consequences of that action. They need not be acceptable in any absolute sense. Those same individuals might choose a riskier course of action (e.g., deciding to pass a slow car), if it brought a compensating benefit. Or, they might choose a less risky course of action (e.g., postponing a trip home until well after the bars close), if that could be done at reasonable cost. A level of risk that is acceptable for one activity might seem horrendously high or wonderfully low in other contexts. In ordinary discourse, it is so easy to lose the essential context of decisions that the term "acceptable level of risk" might best be avoided.10

In this light, a technology should be acceptable to an individual if it creates an acceptable balance of personal risks and benefits. If a technology is acceptable for each member of society, then it should be satisfactory to society as a whole. One might call the risks of that technology societally acceptable (considering its benefits), just as one might call its benefits societally acceptable (considering its risks). A focus on action may produce the best language, that of a societally acceptable technology. This is the definition being advocated here: A technology has a societally acceptable level of risk if its benefits outweigh its risks for every member of society.11

The ethical core of this proposal may be seen most sharply by contrasting it with the utilitarianism of approaches that look at the total benefits accruing to a society from a technology, when judging the acceptability of its risks. A rough method for doing so is to perform a cost-benefit analysis, summarizing economic measures of a technology's total benefits and total costs (including the risks that it imposes). A central ethical assumption of many such analyses is that one should look at the overall balance of consequences for society, while ignoring the balance actually experienced by individuals. Under this assumption, one would not care if a technology made society as a whole better off, at the price of making some of its members miserable. Nor would one care if a few people received very large net benefits, while many others had small net losses; or, if many people had small net benefits, while imposing large net losses on a few (e.g., those living near a landfill that accepts hazardous wastes from a large area).12

The rationale for this indifference to the fate of individuals is often some variant on the potential Pareto improvement principle. It holds that an action is acceptable if its excess of benefits over risks is sufficiently great that those who "win" from the action could compensate losers. However, they need not do so. The losers in these transactions may not know, much less be persuaded by, the efficiency arguments supporting this principle. Nor may they see themselves as winners often enough, in the set of decisions resolved by these procedures, to overlook the apparent injustices of a particular decision. Rather than trusting to any long run, they may want to be compensated in every transaction. It would take only a mildly cynical view of how society distributes its wealth to justify a fear of routinely getting the short end of the stick.13

Concern for the fate of individuals is embodied in regulations, like those reviewed by EPA, that specify acceptable levels of individual risk. For example, the maximum risk to someone living 70 years at a plant boundary might be set at one chance in a million of dying from a particular kind of emission. These regulations do not, however, invoke the benefits to these individuals as a concern in setting risk levels. Conceivably, some notion of benefit may underlie the standards. However, as long as they are not mentioned specifically, one cannot evaluate the appropriateness of the tradeoffs created by these standards. Indeed, the standards seem to deny the existence of tradeoffs. They do not distinguish, for example, between the situation of a middle-aged worker at the plant, who voluntarily lives near the gate to use cheap housing and avoid a nerve-racking commute, and that of a child whose parents could not afford to move when the plant set up shop next door.

No reasonable individual would want his or her personal life to be governed by a rigid acceptable level of risk. Nor should a reasonable society want a single level of risk to govern all technologies, regardless of their other features, including the benefits that they bring. Furthermore, reasonable individuals should not want government to take such inflexible actions on their behalf. It is not logically defensible to set a single level of acceptable risk for all technologies, unless a principled decision has been made to ignore all other factors. Had EPA's survey found that agencies make the same demands of diverse technologies, it would have provided a serious indictment of our regulatory processes.

Thus, again, the present proposal is that a technology is acceptable if it creates acceptable risk-benefit tradeoffs for every member of society. This criterion is advanced as a general political-ethical principle of the sort that would be endorsed by most citizens as a fundamental regulatory philosophy. It allows risks to be balanced by benefits, but protects individuals in cases where the greatest good for the greatest number might come at their expense.

If this proposal appears ethically sound, then the critical question becomes whether it can be implemented. Doing so would require: a fuller definition of "acceptable tradeoffs," a credible means for measuring those tradeoffs and an orderly procedure for applying them to evaluating individual technologies. Subsequent sections discuss these topics in turn, suggesting a work plan for developing practical standards from this conceptual proposal. A final section considers how implementing that plan would affect various stakeholders: industries, regulatory agencies, members of the public, and public interest organizations. It concludes that putting the fate of individual citizens at the center of the regulatory process may actually make life easier for many technologies.

As mentioned, in a democratic society, analysis cannot replace process, only inform it. However, a sound analysis can create a disputable presumption regarding what the outcome of that process should be. A technology that failed this test would bear an extra burden, either to create better tradeoffs for the individuals it affects or to demonstrate why it deserves special dispensation. Citizens opposed to a technology that had passed this test would bear an extra burden of proof to argue why it should pass more rigorous

standards. People opposed to an analytical result will, naturally, criticize its technical details. However, that should be much less frustrating than trying to get the details right for an analysis that itself made no sense.

The goal of this proposal is not to enshrine and defend a single absolutist principle. Rather, it attempts to create a workable compromise. It qualifies and elaborates the core concept in ways that preserve and refine its basic thrust, in the hopes that it will resolve many issues (to the point where they do not seem worthy of debate) and focus debate on the others. It aims at fewer, but better conflicts.

Defining Acceptable Technologies

Conceptually, the most straightforward approach to determining the acceptability of a particular technology is simply asking all relevant citizens whether they are satisfied with how it affects them personally. The present section analyzes the reasons why such direct assessment is not viable. It argues, in effect, that, although the present proposal is for the people, it cannot be by the people -- at least not in this fully democratic sense. This discussion leads to a refined statement of the acceptable technology principle.14

Sampling Limitations

Vast numbers of people are exposed to at least some risk from many technologies. Indeed, where atmospheric distribution of toxins is possible, every person on the planet may be exposed. Given the intensive interviewing needed to elicit informed judgments regarding risk-benefit tradeoffs,15 it is impossible to ask everyone about everything. Thus, the most that can be elicited is general guidance regarding the kinds of tradeoffs most people would accept -- were they asked in a way that allowed them to understand the risks, the benefits and their own attitudes toward the tradeoffs.

Such direct evidence of concern is also the least that can be obtained if public welfare is to be at the center of regulatory processes. For example, one cannot rely on risk professionals' speculations regarding "what the public wants." The limits to expert opinions can be seen in disagreements among them about that.16 Professionals contact the public so irregularly, and their life experiences are so different, that they cannot claim accurate knowledge. Also, there is too much opportunity to report "what I think I would want were I in the public's place" or even "what I would like the public to want." Analogous criticisms limit the value of other potential stand-ins for genuine citizens (e.g., public interest advocates, elected representatives, pundits), although citizens might wish to consider such opinions when deciding which tradeoffs are right for them.

Thus, acceptable tradeoffs must be ones that citizens endorse in principle. They cannot want to be asked to evaluate their fate at the hands of every single technology, much less every change in its operating procedures. Being forced to have an opinion on every problem would mean being denied the opportunity to have articulated opinions on any problem.17

Although one might, in principle, ask every citizen about those general standards, a properly chosen sample should provide estimates of any desired precision. Moreover, the queried sample deserves the chance to develop thoughtful positions on these fateful issues and will need more opportunities for reflection than is possible with conventional survey research.18 Thus, sample size might be sacrificed for measurement accuracy, securing fewer people, but ones who really have a basis for their opinion -- not unlike the situation with common law juries.

Complexity

The impossibility of asking everyone about everything reflects not just the number of potential decisions but also their complexity. For example, the money saved by not reducing a risk typically goes directly to the risk's producer. Nonetheless, other individuals, including those exposed to the risk, may receive indirect benefits as those cost savings pass through the economy. When federal government facilities operate at riskier levels, then taxes might be lower for all citizens, including those who bear the additional risks. Similarly, allowing a privately owned plant to operate more riskily might encourage it to remain in a community, leave it with more capital for local investment and encourage more generous wages. The ripple effects of such actions might benefit risk-bearers who are neither employees nor stock owners. Conversely, technologies should also be held accountable for indirect costs created by their risks. Some may be hard to measure or of uncertain relevance;19 e.g., the anxiety caused by concern over a landfill can have real health effects, even if rooted in misunderstandings.20

Identifying all of these consequences, much less quantifying them, is not work for the timid. A "classic" example of these difficulties might be found in the controversy over the direct and indirect risks of energy systems, prompted by Inhaber's analysis.21 Individuals might at best hope to understand the full set of personal effects for a handful of risky technologies for which they had a particular interest. As a result, it is a proper regulatory function to analyze the risks and benefits that a technology creates for individual citizens, then subject those summary measures of acceptability to the general standards produced by representative groups of ordinary citizens.22

Strategic Responses

Were they asked to judge the tradeoffs associated with a specific technology, properly informed individuals would realize that its fate hinged on their consent. If they sought to get the best possible deal for themselves, then they should exploit this position and demand more benefits than they would ordinarily view as constituting adequate compensation. Indeed, there would be no constraints on their demands beyond a technology's ability to pay. Some people might like the idea of stripping industries of all but the minimum profits needed to be viable. Yet, doing so would involve a greater shift of political power than is likely within any current regulatory system.23 The present proposal has the more modest goal of giving legal standing to the welfare of all citizens, including those whose lack of political standing might otherwise allow the imposition of unacceptable tradeoffs.

In this approach, people are not represented directly, but through their values, namely, the values that they would express in a situation where they could neither exploit some artificial veto power nor be exploited by coercive social arrangements. Such a standard would imply some surrender of absolute sovereignty on both sides were it implemented rigidly. A theory for justifying that restraint on individual choice is that an orderly society needs the limited right to impose risks in return for due compensation, just as it needs the limited right to secure property for the public good (e.g., road building). With declarations of eminent domain, the property is a physical object and the compensation is determined primarily by market value. Here, the "property" is the degree of personal safety that is lost through exposure to a technology. Proper compensation is the level of benefit that people would ordinarily consider to offset risks like those of the technology -- absent any advantages or disadvantages in bargaining position.

Individual Differences

If asked, different people might accept very different tradeoffs. Some may dislike risks to health and safety so much that they demand enormous compensation in return for any exposure. They do this in their own lives, and they expect the same treatment for risks from technological sources. Other people may be so indifferent to such risks that they require relatively little compensation. In their own lives, they do little to reduce risks, even ones whose benefits are minimal. Technological risks bother them equally little.24 It would be hard for a single regulatory policy (or a single configuration of a regulated technology) to satisfy individuals at both extremes. Risk-avoiding individuals would be aghast at the uncompensated riskiness of a technology that satisfied their risk-indifferent counterparts. The latter might be bemused at the resources "wasted" in needless risk reduction. They might be angry if they believed that some of those resources might otherwise come their way.

Thus, a third "compromise" of the ideal of using each affected citizen's values for each situation is needed: Rather than having to satisfy every possible set of values, a technology should be required to produce acceptable sets of consequences for individuals having "reasonable" values. This criterion is analogous to the reasonable person standard, a routine feature of legal proceedings. It is meant to exclude those who fall well outside the normal range. On the one hand, a technology would not have to satisfy individuals who would do almost anything to avoid the sort of risks that it creates. On the other hand, a technology would get no credit for satisfying individuals who care little about self protection or actually enjoy risk exposure.25

Like the other principles advanced in this proposal, if this one is accepted, then work could begin on its implementation (discussed in greater, but still partial detail, in the following section). One possible operational definition of "unreasonable risk avoidance" is willingly taking actions that create risks greater than the ones that they are meant to avoid (e.g., risking malnutrition in order to avoid foodstuffs with minimal pesticide residues). One possible operational definition of unreasonable risk acceptance is routinely passing up low- (or no-) cost opportunities to reduce risks. A fuller implementation might also explore creating a distribution of individuals in terms of their degrees of risk-

aversiveness, then truncate symmetrically at some extreme fractiles. Again, the focus in this essay is on developing a proposal worthy of such detail. The act of defining "reasonable values" means imposing a general societal standard on the individual desires that societally regulated technologies must meet. Some individuals will be told, in effect, that they are not entitled to as much compensation for risk as they usually demand, while others will be getting more than they would ordinarily expect. It is an empirical question whether the same citizens will prove to have unreasonable values in case after case -- or whether different people will prove most averse to different risks.26

Legitimacy of Evaluative Criteria

In its Survey of Risks, EPA listed several conditioning factors that might affect judgments of acceptability. Some of these seem to have been drawn from psychologists' "psychometric" studies of risk, initiated by Fischhoff, Slovic, Lichtenstein, Read and Combs.27 These studies have found that laypeople want higher levels of safety from technologies whose risks have certain qualitative properties, such as being unfamiliar, evoking a feeling of dread and being perceived as poorly understood by science. For example, a technology whose risks were imposed involuntarily would have to provide greater benefits than a technology with the same amount of risk, but whose adoption was voluntary.28 Because technologies vary on these factors, people would not find any single risk level acceptable for all technologies.29

Whether such double standards ought to be imposed on technologies, just because the public wants them, is a matter of regulatory philosophy. For example, one might want greater benefits from technologies that evoke a feeling of dread, in order to compensate its citizens for the attendant loss in quality of life. However, one might also believe that a technology is not responsible for how people feel about it, even when those feelings are based on accurate risk perceptions. One might want greater benefits from technologies whose risks are poorly understood by science in order to encourage better research or to create a reserve for unpleasant surprises. However, one might also want to be neutral toward uncertainty, in order to avoid discouraging new technologies.

The organic or enabling legislation of an agency may, however, have no place for some of these considerations. That is, if left to their own devices, citizens might base their acceptability judgments on factors that have no legal relevance. As a result, citizens asked to evaluate tradeoffs would have to be focused on factors they are allowed to consider by an agency empowered to determine which attributes of risk and benefit are legitimate bases for public policy. Beyond that, it would be required to let representative citizens determine what weight, if any, should be given to each.30

Summary

The acceptability of a technology should depend on the acceptability of its consequences for individual citizens. However, for both practical and philosophical reasons, that determination cannot be left to those citizens. There are too many issues, of too great complexity, for citizens to be able to identify their own best interests regarding every

technology that poses some risk to them. Even if individuals knew their own minds, society does not have the resources to solicit opinions from everyone in every case. Moreover, individuals would have every incentive to demand exaggerated compensation, exploiting the need to secure their personal acceptance of a technology. Indeed, the best informed individuals might also be the most unreasonable. Even individuals who do not respond strategically may have unreasonable demands for protection or unreasonable willingness to accept risks. Finally, judgments that are not out of the ordinary may still reflect concerns that are not appropriate bases for regulatory policy.

These considerations lead to this principle:

A technology is acceptable if it creates an acceptable set of consequences for every member of society. Compliance should be determined by applying a general evaluative standard to the best available estimates of the technology's consequences. That standard should express the values of individuals with reasonable attitudes towards risk, constrained to focus on legally relevant consequences, and allowed to develop well-articulated positions.

The next section elaborates on procedures that could be used for implementation. They require basic research to determine general rules of acceptability (with input from representative citizens and within constraints set by regulations), followed by applied research to apply them to specific circumstances. Such an objective determination of subjective values is needed to protect individuals from being exploited by society and society from being coerced by individuals.

Determining Acceptable Tradeoffs

Implementing a regulatory principle has two steps: developing explicit general rules, thereby defining acceptable performance; and applying those rules to specific technologies, thereby determining their fate. All political-ethical-value questions should be resolved in the first step, so that the second involves only technical application of rules. The first requires political judgment to determine what kinds of tradeoffs are acceptable; the second requires scientific judgment to estimate the risks and benefits of particular technologies with enough precision to determine whether it meets the standard. The first step calls primarily for input from the social sciences, for measuring citizens' general attitudes toward risk-benefit tradeoffs.31 The second calls for inputs from various sciences, for measuring specific risks and benefits.32

As with any practical procedures in a complex world, these will require compromises to be implemented. The problems associated with risk and benefit assessment are well known and debated.33 They will not be repeated here, except to note that they often produce estimates of a variety of consequences (both good and bad), ranging over many orders of magnitude (from the best to the worst), and are often surrounded by considerable uncertainty (regarding both which measures and which models to use). The general rules must apply to tradeoffs among those kinds of outputs.

What follows is a conceptual analysis of how the development of such standards could be organized. It proposes a procedure with three stages: screening, balancing, and adjusting. The screening stage establishes whether, for regulatory purposes, an individual is considered to be exposed to risk from a technology (and would require some compensating benefit). The balancing stage identifies acceptable tradeoffs for exposed individuals. The adjusting stage incorporates additional factors needed to ensure a credible regulatory process, beyond what can be captured in summaries of risks and benefits.

A Measurement Philosophy

Two basic ways to get at people's values are observing their behavior (revealed preferences) and asking them (expressed preferences). As discussed below, the first method is limited unless one can identify the perceptions and constraints that underlie an action. In many cases, it will be extremely difficult to identify any action that clearly reflects many critical tradeoffs.34

As for the second, in principle, one can ask about anything. In practice, though, the fact that we have questions need not mean that our informants have answers. It is difficult to formulate precise value questions, much less render them comprehensible and help people work through the implications of their own preferences.35 Although the present proposal does no more than require directly facing questions implicit in many risk decisions, these questions often make us uneasy. Thus, it seems unrealistic to rely on standard survey methodology, with its dispassionate interviewers presenting questions in a manner designed to avoid any possible influence (or reactivity). Adopting that stance with respondents who lack articulated views means capturing fragmentary opinions and presenting them as deeply held true values. A more appropriate strategy is to work with respondents, helping them to understand issues and develop stable positions. It means striving to balance biases, rather than trying to avoid them altogether. Although unconventional in survey research, such a philosophy underlies decision analysis.36 Its procedures were created for situations with complex consequences and stakes sufficiently high to motivate individual involvement. Surveys are sometimes depicted as mock elections (opinion polls). The procedure required here might better be seen as creating mock commissions or juries, with a random sample of citizens impaneled to work things through on behalf of their peers.

Interviews would, logically, include an element of revealed preferences, and would call upon people to reflect on their own prior behavior and the reasons motivating it. Regarding risks that had apparently been ignored, they could be, e.g., asked: Did they not care? Did they have accurate perceptions? Would they make the same decision again? Did they even think about their actions?

Although daunting, such elicitation need not be perfect. Applying the acceptable-technology principle will lead to the same regulatory decision for all tradeoffs within some range. If an elicitation procedure shows that people's tradeoffs lie within that range,

then that is all that would be needed. The measurement techniques best suited to this task seem to be those of decision analysis.37

Screening

Clearly, people do not pay attention to all risks in their lives, especially smaller ones. This has prompted many proposals for inferring a de minimis level of risk.38 According to these proposals, risks below some level can be ignored when technologies are regulated.

Unfortunately, such proposals fail to ask (or at least determine) whether people really do not care about risks they seemingly ignore. Do people even know they exist? If so, do they really know how great they are? How much thought have they given to the issues? If particular risks seem unacceptable, do people have any avenue, or energy, for expressing concerns? If they accept a technology, is it because the risks are negligible, or because there are compensating benefits? How accurately do they perceive any benefits? Without answers to these questions, no clear conclusions can be drawn from observed behavior.

Once the potential of revealed preference analyses has been exhausted, expressed preferences would be explored. The critical question for standard setting is: What risks are people willing to ignore, so that no compensating benefit is required, if they are exposed to risks at that level? This judgment should not consider the transaction costs of either evaluating risks or collecting compensation. The agency implementing the general standard would handle both. It would commission (or review) the risk analyses. It would exact and allocate any needed compensation, beyond what would occur naturally.39

A successful screening procedure could dramatically reduce the number of individuals whose welfare needs to be considered. Such success might seem improbable in light of the observation that citizens often seem very agitated by risks that many experts view as very small. That suggests that no risk is so small that no compensation is required. The problem with this inference is its simplistic interpretation of citizens' behavior. They may not accept the experts' claims. Or, they may object to the people and the political process managing the risk. That is, they may feel that they are losing rights and respect, but be constrained to talk about risks. Experts may prefer to call the public stupid rather than to admit that they have treated it high handedly.

Balancing

Whenever a technology poses non-negligible risks to an individual, it would then have to pass a risk-benefit test, showing that its consequences are acceptable, as judged by the general standard set for reasonable individuals. These standards can be developed with the same kinds of procedures as are available for screening decisions:

* a. observing the preferences that may have been revealed in past decisions,

* b. asking people in the abstract what tradeoffs they deem acceptable,

* c. asking people to evaluate hypothetical situations that embed abstract tradeoffs in concrete examples, and

* d. asking people to review their own previous decisions, clarifying the tradeoffs that those choices were meant to embody.

Such research would need to provide not only its best estimate of these value judgments, but also an assessment of its own definitiveness, suitable for sensitivity analyses.

In reviewing past decisions, as in procedures a and d above, the greatest credence would be given to cases fulfilling the conditions of informed consent; that is, where the decision-making process was well-informed, thoughtful, and uncoerced. Where people were uninformed or misinformed (regarding risks or benefits), analysts must reconstruct the decisions that people thought they were making. Where people were unable to make thoughtful choices, analysts must divine which choices would have emerged under more favorable circumstances. Barriers to thoughtfulness include being rushed and simply not knowing how to organize one's work. Although time pressure can complicate decision making, poor choices are often made with all the time in the world.40

Actions, to be analyzed at all for evidence of people's values, need to represent decisions, i.e., they need to reflect choices among alternative courses of action. In the language of risk analysis, they need to be voluntary. Deriving standards from involuntary decisions means interpreting as acceptable whatever tradeoffs people have been forced to accept. It means enshrining the injustices of the past in prescriptions for the future.

Adjusting

No set of general rules will apply equally well in all circumstances. To preserve the public credibility and political viability of an approach, an agency must be able to adjust its determinations in situations having crucial features that are not represented in its general rules. On the other hand, if its work is not to become a patchwork of special pleading, then it must attempt to codify those exceptions in advance. Three examples follow, concerning the sort of systematic exceptions that might be applied to the acceptable-technology principle:

* People want not only to receive attractive deals, but also to feel that they have been treated fairly. Even if a technology provides them with an acceptable risk-benefit tradeoff, people may be dissatisfied, for example, if they feel that the technology's sponsors could have paid them more without impairing the technology's economic viability. They may be dissatisfied if a technology was located in their community simply because it cost less to provide them with an acceptable tradeoff than it would have cost to satisfy residents of a wealthier community (who are accustomed to receiving more compensation for any given level of risk). They may be dissatisfied if they believe that others got a better deal than they did.41 These are predictable, familiar human emotions. Whether they have standing requires a regulatory determination. If they do, then the decisions emerging from the screening and balancing stages might have to be adjusted.

For example, an agency might impose a "poverty premium," demanding higher compensation for risks in poorer communities. If not, then the agency should be explicit about the irrelevance of these issues.42

* Citizens might be dissatisfied if the search for additional safety stopped once a technology had been deemed acceptable -- just as its owners' might be dissatisfied if they were still required to incorporate every new safety device. A compromise adjustment might be to require additional safety measures that passed an explicit cost-effectiveness criterion (e.g., reduced radiation exposure for less than $1,000 per person-rem43). Such a rule might reassure the public that there are incentives for developing safety measures (insofar as the inventor of an efficient risk-reduction method could expect to have it mandated), without imposing an unreasonable burden on industry.

* The acceptable-technology principle is exclusively egoistic. In it, individuals judge the acceptability of a technology solely by the risks and benefits that they personally receive from it. However, people also make sacrifices for the sake of others. For example, the neighbors of a landfill might tolerate somewhat higher risk levels if the alternative was shipping the waste to a developing country, or if they felt that this was their part in a social process that ensured an orderly distribution of risk burdens across the country. Assuming that some underlying order can be discerned, altruistic adjustments might also be incorporated in the standard.44

Finally, of course, after screening and balancing, adjustment would require legal or administrative mechanisms to integrate determinations from this procedure with others that might be legally required.

Summary

The core of any safety determination is a value judgment defining the "acceptability" of risk. In the approach proposed here, that judgment is expressed in terms of a performance standard, specifying risk-benefit tradeoffs that a technology must produce, rather than a technical standard, specifying design and operation details. This requires that performance be evaluated by individual citizens. Once operational, it would involve a set of general rules for: screening cases, to eliminate those where a technology poses a negligible risk; balancing risks and benefits in the remaining cases, thereby characterizing acceptable tradeoffs; and adjusting the balance statement to accommodate additional factors.

Deriving such rules will require detailed analysis, using some combination of the procedures outlined here. Some of these methods rely on what people say about their values, whereas others rely on what people actually do when their values are at stake. None are perfect. All provide some complementary insights, assuming that their strengths and weaknesses are understood. The more satisfactory their implementation, the fewer issues will have to be addressed on an ad hoc basis when the standard is applied. As mentioned, the product of applying this procedure would be the point of departure for political processes, wherein the affected parties struggle over the acceptance of its

recommendations. That is the fate of any regulatory decision. The hope with this proposal is that the ensuing struggles will be fewer and better focused, by virtue of embodying a principle that places individual citizens' welfare at its core, provides industry with a predictable standard and sensible incentives, and anticipates the major exceptions.

Applying the Standard

Compared to existing approaches, a new approach has the advantage of being unsullied by failures and compromises. Yet, it has the disadvantage of having all the hard work of implementation in front of it. Whether further elaboration is warranted depends on its promise. As an aid to appraising its promise, this section considers potential challenges to its practicality and political acceptability.

Practicality

Perhaps the most obvious practical problem with this approach is the apparent need to calculate risks and benefits separately for every individual exposed to non-negligible risks (and to establish the negligibility of the risks to everyone else). Where there are many such individuals, this could require horrendous computation. Making it manageable will require a structural analysis of how risks and benefits are distributed. For example, the estimation process might begin with the individuals bearing the greatest risk (to see if their benefits are commensurate)45 or with the individuals receiving the least benefit (to see if their risks are non-commensurate). One then might look for individuals experiencing intermediate levels of risk who receive unusually low benefits, before considering individuals (or classes of individuals) whose situations bear detailed analysis.

A common situation will be individuals who receive little risk and little benefit from a technology. Very small risks may come from very unlikely worst-case scenarios or from pollutants distributed widely at very low concentrations. Very small benefits may come from diffuse contributions to the overall economy. For example, a factory may ever-so-slightly reduce taxes or the threat of unemployment for individuals living on the other side of town or state. Thus, the acceptability of the technology for the vast majority of people will involve roughly the same risk and benefit estimates. In some cases, applying the general rule will yield so clear-cut a result that no refinements are needed. It may be that the small benefit far outweighs the minimal risk, even for the most averse (but still reasonable) individuals in that class. Or, the technology may be so far out of line that a redesign or compensation plan is needed before proceeding with the analysis.

All regulatory approaches must contend with uncertainties left by even the best risk and benefit assessment methods. To be treated systematically, uncertainties must be summarized quantitatively.46 One can then determine whether a technology is in compliance, to any desired degree of confidence (e.g., can one be 95% certain that the benefits outweigh the risks?). A technology's impacts may be poorly understood in an absolute sense, but still be well enough known to allow a regulatory judgment. The level of confidence demanded would be part of the general standard.

Uncertainties are often particularly large where risks and benefits are particularly small.47 Fortunately for the sake of the analysis, people may demand less precision here than with more consequential technologies. The screening procedure will show some risks to be negligible, whereas others can be justified by any arguable benefits. If so, a rough calculation of risk and benefit might be enough to demonstrate acceptability with adequate confidence. Further work might reveal other shortcuts that are scientifically and ethically acceptable. For example, people might accept replacing person-by-person assessments with class-by-class assessments, especially if the science was better at the aggregate level. Rather than estimating the impacts for each individual residing ten to fifty miles from a factory, one might estimate the total risks and benefits accruing to these individuals, under the assumption that they bear roughly equal shares of each.

Political Acceptability

Ideally, a regulatory proposal would be evaluated solely in terms of the general ethical principles that it embodies, its compliance with legal constraints, and its practicality -- and not in terms of the specific decisions that it will produce.48 To that end, the formulation of this proposal has been motivated by the desire to address the principled objections that can be raised against utilitarian philosophies (which ignore distributional effects) and risk-only philosophies (which ignore all other effects). It was hoped that this appeal would foster the patience needed to work out its technical details.

Realistically, though, people will judge the procedure by how its application will affect them personally. One possible defense against such strategic behavior is to argue that regulatory processes are so complex that it is hard to predict their outcomes. As a result, honesty is the best policy when designing general procedures. If the principles underlying a proposal are sound, then one should trust it to allow the merits of one's position to emerge in specific applications. The following paragraphs discuss what each of three groups of stakeholders might find if it tried to project how this approach would affect its vested interests.

For regulators, a working version of this proposal would make regulatory operations more efficient, by reducing them to the routine application of an accepted rule, and less controversial, by concentrating political-ethical questions in the role-development process. Even failures might be relatively productive, if the logical coherence of the approach made it relatively easy to diagnose their sources. That might make it easier for an agency to press legislators for a clearer mandate. On the other hand, any change in procedure brings disruption, surprises, and the need for a transition period.

For industry (or government, when it sponsors technologies), any predictable, efficient process should reduce costs due to regulatory delays and unpleasant surprises. Having to provide an acceptable balance of consequences for all affected individuals is a rigorous standard. However, it is also one that provides considerable design freedom in achieving compliance (whether by increasing benefits or by decreasing risks). The focus on individuals also offers a potential solution to the recurrent problem of what to do when large numbers of people receive small exposures, without resorting to de minimis

arguments like "they shouldn't mind a little risk" and without having to choose among the competing models for estimating low-exposure risks. People may agree to let a vague chance of a very small risk be balanced by a vague chance of a very small benefit -- if the choice is made within a generally credible procedure.

For citizens, the approach officially places the welfare of individuals at the center of regulatory policy. It offers an explicit set of procedures, open to review at both the standard-setting and standard-application stages. It does, however, undermine the legitimacy of risk-only standards which have sometimes been favored by public interest advocates. Whether those advocates would oppose this proposal should depend on whether they have promoted rigid risk standards primarily as a strategic position, designed to manipulate regulatory processes that are seen to underweight risks to the public. Recognizing both the risks and the benefits of technologies, as proposed here, seems like a reasonable compromise.

Conclusion

Orderly regulation requires well-specified, logically defensible procedures. Without them, regulation is chaotic, unpredictable and frustrating, with little promise of providing either the sort of protection the public desires or the sort of stable environment that industry needs. Within these goals, an approach has been developed that makes the welfare of individual citizens the primary concern of regulatory processes, while still providing industry with a clear, flexible, and sensible set of requirements. A plan is sketched for putting this conceptual proposal into practice. Details are necessarily sketchy and merit elaboration only if the proposal seems practical enough and political enough to offer the possibility of a more orderly and coherent treatment of acceptability -- a task that so far has defied our best efforts, especially efforts attempting to specify a fixed level of "acceptable risk."

Notes

* Support for this research was provided under National Science Foundation Grant SES-8715564. It is gratefully acknowledged, as are the helpful comments of Cindy Atman, Adrian Cohen, Robyn Dawes, Hadi Dowlatabadi, Thomas Field, Gregory Fischer, Lita Furby, Neil King, Eric Males, Jon Merz, Granger Morgan, Tom Smuts, Ola Svenson, and Earle Young, Jr., as well as those of several anonymous reviewers. The views expressed are those of the author.

** Dr. Fischhoff is Professor of Social and Decision Sciences and of Engineering and Public Policy at Carnegie Mellon University. He received his B.S. (Mathematics) from Wayne State University and his M.A. and Ph.D. (Psychology) from Hebrew University, Jerusalem.

1 Howard I. Adler & Alvin M. Weinberg, An Approach to Setting Radiation Standards, 34 Health Phys. 719 (1978); James O. Corbett, Risk Assessment Criteria for Radioactive Waste Disposal, 8 Risk Anal. 575 (1988); Health & Safety Executive, The Tolerability of

Risks from Nuclear Power Stations, (London 1987); William R. Lowrance, Of Acceptable Risk (1976); Paul Milvy, A General Guideline for Management of Risk from Carcinogens, 6 Risk Anal. 67 (1986); Chauncey Starr, Risk Criteria for Nuclear Power Plants: A Pragmatic Proposal, 1 Risk Anal. 113 (1981); and Chauncey Starr, Risk Management Assessment and Acceptability, 5 Risk Anal. 97 (1985).

2 Environmental Protection Agency, National Emission Standards for Hazardous Air Pollutants..., 53 Fed. Reg. 28,495 (1988).

3 Natural Resources Defense Council, Inc. v. U.S. Environmental Protection Agency, 824 F.2d 1146 (D.C. Cir. 1977).

4 55 Fed. Reg., at 28512-513.

5 Nuclear Regulatory Commission, Safety Goals for Nuclear Power Plants (NUREG-0880 1982).

6 See, e.g., Vickie M. Bier, The U.S. Nuclear Regulatory Commission's Safety Goal Policy: A Critical Review, 8 Risk Anal. 563 (1988); Baruch Fischhoff, Acceptable Risk: The Case of Nuclear Power, 2 J. Pol'y Anal. & Mgmt. 559 (1983); J. Michael Griesmeyer & David Okrent, Risk Management and Decision Rules for Light Water Reactors, 1 Risk Anal. 121 (1981); Kenneth A. Solomon et al., An Evaluation of Alternative Safety Criteria for Nuclear Power Plants, 5 Risk Anal. 209 (1985).

7 Survey of Risks (Dkt. No. OAQPS 79-3. Part I, Dkt. item X-B-1).

8 53 Fed. Reg., at 28,513.

9 See, e.g., Daniel J. Fiorino, Regulatory Negotiation as a Policy Process, 48 Pub. Adm. Rev. 764 (1988); Daniel J. Fiorino, Citizen Participation and Environmental Risk, 15 Sci. Tech. & Human Values 226 (1990); Sheldon Krimsky & Alonzo Plough, Environmental Hazards: Communicating Risks as a Social Process (1988); Controversy: Politics of Technical Decisions (Dorothy Nelkin ed. 1978); Harry J. Otway & Detlof von Winterfeldt, Beyond Acceptable Risk: On the Social Acceptability of Technologies, 14 Pol'y Sci. 247 (1982); Lillie C. Trimble, What Do Citizens Want in the Siting of Waste Management Facilities? 8 Risk Anal. 375 (1988); and Elaine Vaughn, Individual and Cultural Differences in Adaptation to Environmental Risks, 48 Am. Psych. 673 (1993).

10 The (UK) Health & Safety Executive, supra note 1, uses the term "tolerable" to describe risks that are accepted for the time being, until a more attractive tradeoff can be found. See also, Baruch Fischhoff et al., Acceptable Risk (1981).

11 There is no reason why these "benefits" should be restricted to economic consequences or even noneconomic ones for which putative economic equivalents exist. People could in principle, be compensated by peace of mind, feelings of satisfaction, or reduction of other risks. See, e.g., Baruch Fischhoff & Louis A. Cox., Jr., Conceptual

Framework for Benefit Assessment in Benefits Assessment: The State of the Art (Judith D. Bentkover, Vincent T. Covello & Jeryl Mumpower eds. 1985); Baruch Fischhoff & Lita Furby, Measuring Values: A Conceptual Framework for Interpreting Transactions, 1 J. Risk & Uncert. 147 (1988).

12 The controversial nature of such aggregate analyses may be seen in the conflict between Executive Order 12,291, see, e.g., Fischhoff & Cox, supra note 11 and the Court's opinion in the vinyl chloride case, supra note 3. The former requires cost-benefit analyses, examining the overall impact of significant federal regulatory actions, ignoring which individuals get the costs and which get the benefits. The latter prohibits EPA from considering the cost or feasibility of compliance in setting the acceptable level of risk, thereby ignoring the benefits that less costly operation might bring to society as a whole.

See also Michael S. Baram, Cost-Benefit Analysis: An Inadequate Basis for Health, Safety, and Environmental Regulatory Decision Making, 8 Ecol. L.Q. 473 (1980); John T. Campen, Benefit, Cost & Beyond (1986); David W. Pearce, Cost Benefit Analysis (1983); Elizabeth Stokey & Richard Zeckhauser, A Primer for Policy Analysis (1978).

13 See, e.g., Robert Bullard, Dumping in Dixie: Race, Class, and Environmental Quality (1990) and Commission for Racial Justice, Toxic Wastes and Race in the United States (1987).

14 Whether (and how) public opinion needs to be consulted regarding the adoption of a specific standards is, of course, also a matter of administrative procedure and law. For example, the Court's opinion of the Vinyl Chloride case appears to call for the adoption of a standard that the general public would endorse, were it possible to solicit its collective opinion in a way that ensured full understanding of the standards and their implications. EPA's request for comments on its proposed approaches repeatedly mentions concern for the public's desires. Many regulatory procedures call for public hearings, followed by orderly written responses to questions raised in them. The present proposal is intended to comply with these constraints, and perhaps give them structure. Detailed analyses of particular settings must await a future opportunity.

15 See, e.g., Baruch Fischhoff, Value Elicitation: Is There Anything in There? 46 Am. Psych. 835 (1991) and Ralph L. Keeney & Howard Raiffa, Decisions with Multiple Objectives (1976).

16 See, e.g., National Research Council, Risk Perception and Communication (1989); Paul Slovic, Perception of Risk, 236 Science 280 (1987); Vaughn, supra note 9; and Abraham H. Wandersman & William K. Hallman, Are People Acting Rationally? 48 Am. Psych. 681 (1993).

17 See, e.g., Jacques Ellul, Propaganda (1969).

18 See, e.g., Fischhoff, supra note 15.

19 An issue currently in litigation is whether technologies can be held liable for the existence value of natural resources, that is, the value that people assign to the very existence of, say, the Grand Canyon in a relatively pristine state. If the courts decided that existence value had legal standing, then the threat that a technology posed to the environment could become a risk requiring compensation. Measuring these threats and the values attributed to them would be a significant methodological challenge.

20 See Andrew Baum & India Fleming, Implications of Psychological Research on Stress and Technological Accidents, 48 Am. Psych. 665 (1993).

21 See Herbert Inhaber, Risk with Energy from Conventional and Non-Conventional Sources, 203 Science 718 (1979); see also John H. Herbert, C. Swanson & P. Reddy, A Risky Business, 21(6) Environment 28 (1979).

22 Those general standards might specify, e.g., that a technology should be responsible only for the concern that its risks would generate if they were properly understood.

23 Adopting such a confiscatory policy would also raise the difficult question of how to divide the spoils among those citizens who have preferred strategic responses.

24 See e.g., Risk Taking (J. Frank Yates ed. 1992).

25 Such individuals derive unusual benefit from the risk, meaning that their preferred risk-benefit tradeoffs may not be all that different from those of nonrisk seekers.

26 Note that individuals who are extremely averse to risks need not be extremely sensitive to them (a protected class in some regulations). One could respond more acutely to a given exposure to a toxin, yet still not want particularly large compensation particular probability of such a response.

27 See supra note 7; compare Baruch Fischhoff et al., How Safe Is Safe Enough? A Psychometric Study of Attitudes towards Technological Risks and Benefits, 8 Pol'y Sciences 127 (1978)

28 The idea of looking for a double standard was proposed by Chauncey Starr in Social Benefit versus Technological Risk, 165 Science 1232 (1969). A list of features that might prompt double standards was compiled by Lowrance, supra note 1.

Further studies in this "tradition" are summarized in Robin Gregory & Robert Mendelsohn, Perceived Risk, Dread and Benefits, 13 Risk Anal. 259 (1993); Slovic, supra note 16; and Paul Slovic, Baruch Fischhoff & Sarah Lichtenstein, Behavioral Decision Theory Perspectives on Risk and Safety, 56 Acta Psych. 183 (1984).

29 See, e.g., Baruch Fischhoff, Stephen Watson & Chris Hope, Defining Risk, 17 Pol'y Sciences 123 (1984).

30 Where citizens felt strongly about factors that they could not consider, then the agency might maintain two sets of books, one for legitimate factors and one for all factors. Over time, changes in the regulatory climate might allow the omitted factors to be included, much as environmental effects are gradually being incorporated in national accounts. See, e.g., Environmental Accounting for Sustainable Development (Y. J. Ahmad, S. El Serafy & E. Lutz eds. 1989); Ecological Economics: The Science and Management of Sustainabilty (Robert Constanza ed. 1991); and Robert Solow, An Almost Practical Step Toward Sustainability (1992).

31 The humanities might also pay a critical role in formulating possible tradeoff rules, as inputs to the citizens entrusted with expressing public values (through the best-available social-science procedure).

32 See, e.g., Baruch Fischhoff, Setting Standards: A Systematic Approach to Managing Public Health and Safety Risks, 30 Mgmt. Sci. 823 (1984).

33 See, e.g., Bentkover et al., supra note 11; Silvio O. Funtowicz & Jeremy R. Ravetz, Uncertainty and Quality in Science for Policy (1990); M. Granger Morgan & Max Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis (1990).

A recent interchange can be found in the Center for Risk Analysis discussion of the Office of Management and Budget's critique of risk assessment procedures in the federal government; see Office of Management and Budget, Current Regulatory Issues in Risk Assessment and Risk Management (1990) and Center for Risk Analysis, OMB vs. the Agencies: the Future of Cancer Risk Assessment (1991).

34 See, e.g., Fischhoff & Cox, supra note 11; Fischhoff et al., supra note 27; and Paul Slovic & Baruch Fischhoff, Targeting Risks: Comments on Wilde's "Theory of Risk Homeostasis," 2 Risk Anal. 231 (1983).

35 See, e.g., Fischhoff, supra note 15; Fischhoff & Furby, supra note 11; The Origin of Values (Michael Hechter, Richard E. Mischod, & Lynn Nadel eds. 1993); Robert C. Mitchell & Richard T. Carson, Using Surveys to Value Public Goods: The Contingent Valuation Method (1989).

36 See, e.g., Ronald Howard, On Fates Comparable to Death, 30 Mgmt. Sci. 407 (1984); Howard Raiffa, Decision Analysis (1968); Stephen Watson & Denis Buede, Decision Synthesis (1987); and Detlof von Winterfeldt & Ward Edwards, Decision Analysis and Behavioral Research (1986).

An example of the sort of progress that might be made through a focused effort to develop measurement techniques of the sort that can be found in recent studies of attitudes toward fairness. See, e.g., C. Harvey, Decision Analysis Models for Social Attitudes toward Equity, 21 Mgmt. Sci. 1199 (1985); Daniel Kahneman, Jack Knetsch & Richard Thaler, Fairness as a Constraint on Profit-Seeking: Entitlements in the Market,

76 Am. Econ. Rev. 728 (1986); L. Robin Keller & Rakesh K. Sarin, Equity in Social Risk: Some Empirical Observations, 8 Risk Anal. 135 (1988); Barbara Mellers, Fair Allocations of Salaries and Taxes, 12 J. Exp. Psych.: Human Percept. & Perf. 80 (1986).

37 See supra Howard; Watson & Buede and von Winterfeldt & Edwards.

38 See, e.g., Cyril L. Comar, Risk: A Pragmatic de minimis Approach, 203 Science 319 (1979); Joseph Fiskel, Toward a de minimus Policy in Risk Regulation, 5 Risk Anal. 257 (1985); Health & Safety Executive, supra note 1; Milvy, supra note 1; Jeryl Mumpower, An Analysis of the de minimis Strategy for Risk Management, 6 Risk Anal. 437 (1986); Gerald J. S. Wilde, A Theory of Risk Homeostasis, 2 Risk Anal. 209 (1982).

39 That is, where a technology does not inherently provide enough benefits to compensate a group of citizens, it could make direct transfers to them. Because the acceptability of a technology depends on its net benefits (after transaction costs), those who manage it would be motivated to find the most efficient allocation scheme, for which the agency would be a likely vehicle.

40 See J. Frank Yates, Judgment and Decision Making (1989) and Yates, supra note 24.

41 See, e.g., Lita Furby, Psychology and Justice in Justice in Views from the Social Sciences 153 (Ronald L. Cohen ed. 1986); Kahneman, supra note 26 and Krimsky & Plough, supra note 9. See also recent studies mentioned supra in note 36.

42 E.g., it could say "In setting standards for specific technologies, the agency cannot address issues of economic equality nor can it consider people's jealousy or upset regarding their neighbors' fate -- as long as they have been treated in accordance with our general principles. If those feelings and inequities have any standing, they would have to be addressed within the context of other federal policies, such as income tax rates."

43 Supra note 5.

44 This is one possible way to represent concern for future generations. The obvious alternative would be calculating the consequences for future individuals. As one got very distant in time, both costs and benefits would often (but not always) become vanishingly small, so that the computational load need not be overwhelming.

45 This would be in keeping with EPA's practice of calculating risks to the maximally exposed individual.

46 See, e.g., Funtowicz & Ravetz and Morgan & Henrion, supra note 33.

47 Consider, e.g., the controversies over threshold effects or indirect economic impacts.

48 See, e.g., John Rawls, A Theory of Justice (1971).