September 1996

Comparing Risks Thoughtfully

Adam M. Finkel

Comparing Risks Thoughtfully

Adam M. Finkel[*]

For very good reasons, comparing risk is becoming all the rage, but the practice of comparative risk assessment (CRA) has caused battle lines to be drawn. I believe that CRA is neither as pernicious as its detractors claim nor nearly as useful as its proponents allege — particularly as is often practiced. At a time when agencies are being required to set risk-based priorities and to compare might-be-regulated risks to others with which the public is familiar,[1] I argue that we need better tools for showing when (if ever) such comparisons are edifying. This article aspires to be an example of what risk assessment practitioners have to offer, if and when the "risk wars" cool down enough to allow a sensible alternative either to no analysis or to runaway analysis as an end in itself.

Based on an appendix to one of the last reports to be released by the Congressional Office of Technology Assessment (OTA),[2] this paper has been revised to address broadly why CRA is controversial and to offer suggestions for improving it.

Thesis

Comparing risks is not impossible or immoral, but it is very difficult[3] — more so than either supporters or detractors of the practice seem to realize. Whenever analysts, agencies or interested parties decide to juxtapose different hazards and compare their severity, they invite controversy. Many stakeholders view CRA with

[1] See, e.g., Dalton G. Paxman, Congressional Risk Proposals, 6 Risk 165, 179 (1995).
[2] Risks to Students in School (1995).
[3] See also, e.g., Baruch Fischhoff, Ranking Risks, 6 Risk 191 (1995).

suspicion because it may usurp fundamentally political choices with the authority of "science" or appear to answer questions it cannot reliably resolve. I too take a skeptical view but reject blanket condemnation that has been partly responsible for dissuading researchers from doing studies such as the OTA report this article originally supplemented.

It would be simple, and perhaps sufficient, to conclude that CRA is inevitable and should therefore be done explicitly — and with broad participation from experts and laypeople. I say "inevitable" because, after taking a close look at the various purposes to which risk assessment is put, it seems that *virtually all risk assessment is CRA*. Standard-setting, the major arena of risk assessment, in fact often entails a risk comparison — that of the existing risk to the reduced risk expected following the contemplated intervention (which may be different in kind or affect a different population). The incessant decisions to focus on some risks and ignore or defer others also involve risk comparison, whether the priorities thus obtained arise via conscious design or by reflex. Even the language of risk communication involves comparison, for few if any of us can really understand the magnitude of a risk of "one in X" or "10^{-Y}" in the abstract without making conscious or subconscious comparison to other risks or situations — to ground these probabilities to an experiential reference point, e.g., an annual·risk of one in 1,000,000 is less likely than being struck by lightning, but more likely than being struck by a meteorite.

CRA is more than a "necessary evil;" it is the logical extension of the less formal thought processes individuals and governments rely upon to help them make choices in all areas. Whenever a decision to — or not to — act has consequences, in most cases more information about the nature and extent of those consequences will make for a better decision. Hamlet's observation that "there is nothing either good nor bad, but thinking makes it so"[4] applies to the two major components of CRA: Thinking about what advantages and disadvantages (or "costs and benefits," if you will) that different situations will bring about, and then thinking about how desirable or undesirable these consequences are. The former process tends to be more descriptive than normative, and the latter more a matter of values

---

4   Act II, ii, 259.

than of analysis, but together they comprise what the late Bart Giamatti called "the edifice of belief" that provides the basis for enabling us "to cope with the vast population of decisions we all live in."[5]

Just because an activity like CRA can be both inevitable and beneficial in theory, however, does not mean that it will do more good than harm in practice. The history of CRA has certainly been marked by strident — and well-founded — criticism of its products. Among many observers, especially those who tend to favor strong environmental, health and safety regulation, CRA has achieved notoriety as an instrument for public confusion and disguising value judgments with scientific veneer. This article will examine whether such criticisms are valid and, more importantly, try to make the case that the problems with CRA are not inherent in the method — but are the fault of the methodology of its practitioners and can be remedied. To borrow from Shakespeare again, "The fault, dear Brutus, is not in our stars, But in ourselves, that we are underlings."[6]

The lion's share of this article will discuss two major pitfalls of CRA — one commonly cited by opponents and supporters alike, and one routinely ignored. I then will sketch out a framework for improving CRA to help ameliorate these problems. To foreshadow where this argument is headed, I will use the metaphor of "the keys and the lamppost," and suggest that at present, CRA is an excellent method for finding our lost keys if in fact they lie within the circle of light cast by the lamppost. If the answer is not there, we need to improve CRA to illuminate a broader area.

When and How CRA is Currently Undertaken

Risk comparisons are ubiquitous in the daily lives of individuals and societies. Planning for the national defense, for example, is a constant endeavor of assessing and comparing threats of disparate types in arenas all over the world, and allocating finite resources to minimize the chance that we will be unable to respond quickly to a serious hazard. Similarly, U.S. states (notably Oregon) and those abroad, who are wrestling with the inability to provide high-technology medical

5 A. Bartlett Giamatti, A Free and Ordered Space: The Real World of the University (1988).
6 Julius Caesar, Act I, ii, 134

interventions to all their citizens, have had to compare the costs and efficacies of hundreds of procedures and therapies to determine which should be universally available.

In the more germane area of environmental CRA, I have found it useful to distinguish between two different kinds of comparative activities that differ in motivation as well as in methodology. For want of a better pair of descriptions, these can be called the "small" and "large" versions of CRA. "Small" CRA involves the quantitative side-by-side comparison of single risks. Ten or fifteen years ago, the most well-known (actually, notorious) examples of "small" CRA were the juxtaposition of markedly dissimilar risks, often with one of the pair voluntary and the other the result of involuntary exposure. Such "hang-gliding is riskier than benzene"[7] comparisons fell into disrepute because they were seen by many as manipulative and grounded in numerical sleight-of-hand rather than in a neutral desire to inform and help put risks in perspective.[8] Still, the "didactic" type of CRA where disparate risks are compared has never fallen completely out of favor and may in fact be making a comeback. As recently as April 1994, a televised documentary[9] used as its centerpiece a table of risks purporting to show, among other things, that smoking cigarettes takes several years off the life of the average person while exposure to pesticide residues only reduces life expectancy by 27 days on average. Table 1 presents various different examples of "small" risk comparisons, arranged in a rough and highly subjective ordering beginning with the least useful varieties.[10]

---

7   Bernard Cohen & I-Sing Lee, *A Catalog of Risks*, 36 Health Phys. 707 (1979).

8   *Rothschild's Numerate Arrogance*, 276 Nature 429 (1978) (Editorial, calling such comparisons "the kindergarten of risk").

9   ABC News, April 21, 1994, *Are We Scaring Ourselves to Death?*

10   The purpose of this table is to show the wide variety of types of risk comparisons; the ordering should certainly not be construed as challenging the validity of any ordering previously proposed.

Table 1
Different Kinds of "Small" Risk Comparisons
(in one possible ascending order of legitimacy)

- Incongruous, unrelated, mischaracterized risks: "More people died in (some notorious event) than died at Three Mile Island"
- Unrelated risks, but ones a person might conceivably choose among: "Working at an oil refinery is safer than hang-gliding"
- Related risks from different sources: "The cancer risk of eating one peanut butter sandwich (containing 2 ppb aflatoxin) is larger than that of drinking a glass of water containing one ppb chloroform"
- Synonymous risks from different sources: "You are exposed to more benzo[a]pyrene from a pound of char-broiled steak than you are living for a year next to Acme Industries"
- Various risks competing for some of the same financial and human resources: "The 'Valley of the Barrels' site has the 24th highest score on the list of Superfund sites; the 'Sinatra Swamp' is #424 and is thus of much lower priority"
- Existing risk versus its substitute: "More people would get lung cancer if cars burned MTBE than if they continued to burn gasoline"
- Countervailing effects of the same intervention: "Making cars out of high-tech plastic would cause X fewer cancers from oil-industry and steel-industry emissions, but Y more people would die in highway crashes due to less crashworthy autos"

"Large" CRA is a more recent phenomenon. It involves the comparison of categories of risks and is increasingly being undertaken, both for symbolic and practical purposes, as a national consensus begins to grow that as a society we focus too much attention on certain "lower-risk" categories while failing to address other higher-risk categories. The most prominent examples of "large" CRA have come from EPA. Reports in 1987[11] and 1990[12] both tried to show that if we would set our priorities with a more "rational" risk-based mindset, we could save more lives and provide greater ecological protection without increasing our total environmental budget. For example, perhaps the single most commonly-cited "irrationality" is the disparity between the several billion dollars we spend annually to clean up

---

[11] U.S. Environmental Protection Agency, Unfinished Business: A Comparative Assessment of Environmental Problems (1987).
[12] Science Advisory Board, U.S. Environmental Protection Agency, Reducing Risk: Setting Priorities and Strategies for Environmental Protection (1990).

abandoned hazardous waste sites, which EPA estimates in total probably to cause no more than several hundred excess cancer deaths per year, versus the $100 million (at most) we spend controlling radon gas in buildings, which may cause as many as 20,000 excess cancers annually. According to these figures, we might be able to save lives some 20,000 times more efficiently (on a cost-per-life-saved basis) by transferring dollars from site cleanup to radon mitigation. "Large" CRA has not yet made a substantial impact in the way resources are actually allocated, but the continued prominence of these sorts of comparisons in the mass media may well lead to a "quiet revolution" in our national environmental protection system over the next decade.[13] Indeed, CRA's influence may eventually extend even wider, as momentum seems to be building for comparing environmental "unfinished business" with unaddressed risks *elsewhere* in society, perhaps with the end result that we will avail ourselves of other risk reductions (e.g., the provision of smoke detectors in all public housing units or the expansion of prenatal clinics) with resources freed up by the rollback of some environmental programs.[14]

Having addressed the issue of *when* CRA is performed, it is a bit easier to summarize how it is usually carried out. The overwhelming tendency uniting all of the types of comparisons in Table 1 (as well as many other kinds of "small," and most "large," risk comparisons made between federal risk-reduction programs) is the circumscribed nature of the analysis. Risks differ in many ways, whether they are as unrelated as hang-gliding and benzene exposure are, or as seemingly synonymous as lead in paint versus lead in gasoline. Yet, CRA has become a classic exercise in turning on the proverbial lamppost to illuminate a small (though doubtless important) portion of the entire tableau. *What we call "risk comparison" is generally little more than "fatality comparison"* — from all the ways in which two hazards differ, we report the total number of deaths each may cause as if that were the only factor distinguishing them.[15]

[13] Adam M. Finkel, *Taking Aim at Environmental Risks: Questions of Feasibility and Desirability,* The Geneva Papers on Risk and Insurance, July 1992, at 46.

[14] *See* Stephen Breyer, Breaking the Vicious Circle (1993), *but see* Adam M. Finkel, *A Second Opinion on an Environmental Misdiagnosis: The Risky Prescriptions of Breaking the Vicious Circle,* 3 NYU Env'l L.J. 295 (1995).

[15] To be fair, CRA practitioners sometimes compare individual risk levels rather

Actually, the current state of affairs is a bit worse. More influential "risk comparisons" are in fact based on a single quantifiable attribute of the things being compared, but one that may have nothing to do with risk! For example, many critics of our national effort to control synthetic pesticides ritually invoke the statement made by Bruce Ames that "99.99% (by weight) of all the pesticides in our diet are naturally-occurring."[16] Even if this statistic itself was uncontoversial (which it isn't), it is not a statement about risk, but instead about the relative *weight* of the added versus endogenous chemicals; only if the carcinogenic potency (per unit of mass) of the two classes was exactly or nearly equal would "99.99% all-natural" be relevant to risk. Similarly, statements such as "a person whose [daily] diet consisted of nothing but four and a half pounds of applesauce... would ingest in one year an amount of UDMH equal in weight to the tar inhaled by smoking two filtered cigarettes"[17] were influential in the public controversy and private litigation surrounding Alar. However, because it also ignores the crucial variable of biologic potency, this weight-weight comparison is about as scientific (and not as candid) as telling someone that the amount of salt he is putting on his potato would be enough to wipe out a city — if it happened to be plutonium instead of salt!

The remainder of this article will deal largely with the question of why such "one-dimensional" comparisons — assuming the one dimension compared is at least a measure of risk — are a poor foundation for risk communication, decision-making and priority-setting. To make this point, I need to convince the reader of two assertions: (1) There are other legitimate ways to measure the consequences of a hazardous exposure or activity beyond simply the expected number of fatalities; and (2) quantifiable consequences alone are only one of the legitimate lenses through which to view threats to health, safety and the environment.

than population consequence figures. However, in many cases little new information is revealed by adding this dimension; it often simply represents the population-wide consequence divided by the estimated number of persons exposed to each risk, i.e., the average individual risk.

16 Bruce N. Ames, Margie Profet, & Lois S. Gold, *Dietary Pesticides (99.99% All Natural)*, 87 Proc. National Acad. Sci. 7777 (1990).

17 Joseph D. Rosen, *Much Ado About Alar*, Issues in Science and Technology, Fall 1990, at 85.

The Irony of Risk Comparisons

Before turning to ways to improve CRA, it must be acknowledged that improvement may be futile because of the widespread preconception that no method can compare disparate risks sensibly. The politics of CRA have remained stable for a number of years: The "camp" of proponents (consisting largely of academics, industry representatives and some government officials) believes CRA is an underutilized tool that could help us make better decisions, while the "camp" of detractors (consisting largely of environmental advocates and members of the public-at-large) is profoundly suspicious of CRA and believes it cannot be performed thoughtfully enough to overcome its inherent disadvantages. The former group is generally satisfied with existing methodology for comparing risks. It stands to reason that those who believe straightforward comparisons of the likely magnitude of different risks are informative would want the comparison process itself to be simple and standardized. Members of the latter group appear to be primarily concerned about the tendency of CRA to oversimplify and mislead the public by reporting subjective opinion about the relative importance of risks as if it were an wholly objective, empirical observation. Since practitioners of CRA have not made any great strides in addressing this problem, the skeptics' belief is understandable. Because of the nature of this political landscape, there currently is little if any clamor for reforming CRA methodology; the debate tends to be polarized between those who want CRA expanded and those who argue that CRA should be scrapped.

The intensity of the opposition to CRA is actually quite ironic. The feature of CRA that arouses the most criticism at present is in my view not serious, whereas other attributes of current CRA methodology that actually make it more precarious receive little or no attention. Thus, to return to the thesis of this article, critics of CRA may be right — but for the wrong reasons.

By far the most commonly-criticized attribute of CRA is its reliance on juxtaposing situations that allegedly are incommensurable. Opposition to CRA often begins and ends with the put-down that "you can't compare apples and oranges," as if that accusation was so serious that nothing else about CRA mattered. But the impotence of this accusation is reasonably easy to document. The maxim about apples

and oranges is a truism, as opposed to a truth, because it is not generally valid either literally or metaphorically.[18]

Table 2
How to Compare Apples and Oranges

Step I: Disaggregation (Identifying Common Attributes that Matter)
- Taste
- Calories
- Price
- Convenience
- Vitamin C content?
- Vitamin K content?

Step II: Normative Valuation (Gauging Preferences for Each Attribute)
- "More is better" (e.g., taste)
- "More is worse" (e.g., price)
- Inter-individual variation in preferences over the same attribute
- ("Ultra Slim-Fast" and "Quick Weight Gain" both sell)

Step III: Descriptive Valuation (Assessing Each Option and Attribute)
- Which is cheaper?
- Which tastes better?
- Which must be peeled/which washed?
- How do you feel about the Florida Citrus Commission
- (national spokespersons; Anita Bryant, Rush Limbaugh)?

Step IV: Aggregation (Combining Pros and Cons for Each Attribute)

Comparison Example

Attribute	Apple (A)	Orange(O)	Importance	"Subtotal"
Taste	++	+	5	A
Price	$0.79	$0.69	4	O
Calories	60 kcal	70 kcal	3	A
Convenience	++	+	2	A
Politics	0	-	1	A

Rough multiplication and summation of the third and fourth columns gives 11A, 4O; buy the apple!
Even if the scores for one dimension (say, calories) were too uncertain to assign a definite preference to A, the total score would still be 8A, 4O.

Millions of times each day, consumers at their local supermarket assuredly do compare apples and oranges, just as we all do in a less literal sense when making myriad other choices. We compare highly dissimilar states such as marrying or remaining single, going to law or business school, or buying a small house in a safe neighborhood or a larger one in a less safe neighborhood, by a conceptually simple

---

18 Other truisms, such as "a watched pot never boils," may have real value if not taken literally but are still not literally true. (Your gaze cannot affect the transfer of heat from the burner to the water.)

cognitive process. As Table 2 shows, when we choose between two commodities, risks or life prospects, we really perform a mental exercise that involves breaking each option down into its component attributes, making a set of descriptive judgments (How much of each attribute does each situation embody?) followed by corresponding normative judgments (How much do I value each attribute?), and then reconstituting each option by combining the pros and cons contributed by each judgment. So, if we have to choose between an exciting but unreliable life partner and a dull but sturdy one, we can. This implies that we can discriminate between apples and oranges or between smoking and living near a coke oven at least as readily.

Table 3
Ranking Risk Comparisons[19]

First-rank risk comparisons
    1. Comparisons of the same risk at different times.
    2. Comparisons with a standard.
    3. Comparisons with different estimates of the same risk.
Second-rank risk comparisons (less desirable)
    4. Comparisons of the risk of doing and not doing something.
    5. Comparisons of alternative solutions to the same problem.
    6. Comparisons with the same risk experienced at other places.
Third-rank risk comparisons (even less desirable)
    7. Comparisons of average risk with peak risk at a particular time or place.
    8. Comparisons of the risk from one source of a particular adverse effect with the risk from all sources of that same adverse effect.
Fourth-rank risk comparisons (marginally acceptable)
    9. Comparisons of risk with cost, or of cost/risk ratio with cost/risk ratio.
    10. Comparisons of risk with benefits.
    11. Comparisons of occupational with environmental risks.
    12. Comparisons with other risks from the same source, such as the same facility of the same risk agents.
    13. Comparisons with other specific causes of the same disease, illness or injury.
Fifth-rank risk comparisons (rarely acceptable — use extreme caution)
    14. Comparisons of unrelated risks.

---

[19] Table adapted from Roth et al., *What Do We Know about Making Risk Comparisons?* 10 Risk Anal. 375, 376 (1990) — based in turn on Vincent T. Covello, Peter M. Sandman & Paul Slovic, Risk Communication, Risk Statistics and Risk Comparisons: A Manual of Plant Managers (Chemical Mfrs. Assn. 1988).

That disparate circumstances are commensurable when push comes to shove may explain some of the most provocative new empirical research about how laypeople actually regard various types of risk comparisons. Recently, when two groups of researchers empirically tested the most widely-accepted predictions about how laypeople were supposed to react to various kinds of risk comparisons, responses either did not support or flatly contradicted the thesis that the more dissimilar the comparison, the less acceptable and more aggravating the recipients would find it.[20] Table 3 represents an initial ranking by Covello et al. based on limited experience.

For example, those surveyed by the Roth group generally regarded a hypothetical comparison of two estimates of the same pollutant risk (a type of comparison estimated as very acceptable) as *less* reassuring, informative and trust-engendering than a comparison of the pollutant risk with the risk of lightning, hurricanes and insect bites (estimated as "fifth rank" or "rarely acceptable").[21]

The first giant step towards reforming CRA, therefore, is to recognize that the problem is *how* risks differ, not merely *that* they differ. In other words, apples and oranges are not interchangeable, but neither are they totally alien — they are just different, and different in ways we can make sense of and value for ourselves. *The unmet challenge of CRA is to describe disparate risks in rich, informative and non-manipulative ways.*

This conceptual challenge cannot be met, however, unless a technical hurdle is overcome — the damaging tendency to compare risks that are both uncertain and variable via misleading point estimates of each risk. This is what I mean by "the irony of risk comparisons" — the fact that critics of CRA focus on the incommensurability "problem" while practitioners continue to deflect the more serious problem of incorporating considerations of uncertainty and variability into risk

---

[20] Roth et al., *supra* and Paul Slovic, Nancy Kraus & Vincent T. Covello, *Comment: What Should We Know About Making Risk Comparisons?* 10 Risk Anal. 389 (1990).

[21] I do not presume here to judge which of the two papers cited *supra* in note 18 more correctly captures the public's view about the relative acceptability of various types of risk comparisons. I only suggest that the emphasis of Slovic et al., *supra* note 20, on the dissimilarity of the things being compared may be a blind alley if in fact the public recognizes the limited applicability of the "apples and oranges" maxim.

comparison.[22] The next section, with Appendix A, briefly introduces uncertainty and variability in risk estimation and comparison.

Uncertainty and Variability

Uncertainty and variability are both hard to define precisely and hard to keep conceptually separate where appropriate, but both are inextricably part of estimating the magnitude of risk, and even of assessing many of the other less-quantitative attributes enumerated in the following section. Brief definitional material may help.[23]

- Uncertainty is a property of our ability to observe or understand a system; it prevents us from knowing the "true value" of some quantity because we cannot measure it precisely, observe enough information to pin down its behavior, or specify the underlying physical or biological model that would allow us to predict its value given some information. Variability is a property of the system itself; it manifests itself as a multiplicity of "true values" because quantities of interest change over time, across spatial areas or among individual people.
- Applied to a risk estimate, acknowledging uncertainty would mean saying that "the risk to the average person is somewhere between 10^{-A} and 10^{-B}; we cannot say for sure which end of this range (or any value in between) is the true value."[24] Acknowledging variability means saying that "we know the risk to the average person is exactly 10^{-C}, but because people's exposures vary, we can only say that your risk is somewhere between 10^{-D} and 10^{-E}."

---

[22] A further irony is that, in the past several years, observers have begun to recognize that individual risks cannot be properly characterized via point estimates. Yet, this realization has not yet found its way into the debate about CRA even though it is *harder* to compare uncertain risks than it is to make a reasonable attempt to address an uncertain risk in isolation; Adam M. Finkel, *Towards Less Misleading Comparisons of Uncertain Risks: The Example of Alar and Aflatoxin*, 103 Env'l Health Persp. 376 (1995). This assertion flies in the face of much rhetoric about the relative ease of comparing risks (allegedly, putting two or more risks on a relative rather than an absolute scale sidesteps problems of modeling and scale), but a bit more thought should illuminate the more important point that uncertainty compounds with the number of items being compared. Mathematically, the uncertainty in the quotient of two quantities is larger than the uncertainty in each item singly.

[23] For an elaboration, *see* Committee on Risk Assessment of Hazardous Air Pollutants, National Research Council, Science and Judgment in Risk Assessment, Chs. 9–11 (1994).

[24] Equivalently, that "the number of expected fatalities in a population of size P is somewhere between P x 10^{-A} and P x 10^{-B}, not exactly one or the other."

- Since most risk estimates are both uncertain and variable, the range of estimates that would account for both factors will generally be broader than either range assessed separately. For example, a risk assessor might summarize both phenomena by saying "there is a 5% chance that your risk is as high as X" (where X is related to the upper bound given uncertainty on the estimate that applies to a person at the upper end of the distribution of variability) "and a 5% chance your risk is as low as Y."
- Uncertainty can be reduced through additional research or data collection; variability cannot be reduced, only better understood, through investigation of the relevant temporal, spatial and inter-individual factors.
- However, both uncertainty and variability can be "managed" by homing in on a single estimate from their respective distributions. An uncertain risk can be summarized via a central estimate, a lower bound or an upper bound;[25] this value judgment addresses the issue of whether (and to what extent) to strive to be "safe rather than sorry." A variable risk can be simplified by choosing to address a portion of the population or set of temporal/spatial conditions; this may result in value judgments about "Who (or when or where) will be safe, and who will be sorry?"
- When comparing risks, uncertainties or variabilities sometimes cancel (e.g., inter-individual variability is not a problem when assessing the relative risks of two hazards to the same person) but sometimes reinforce each other (e.g., if Risk A causes somewhere between ten and 100 deaths, and Risk B between five and 50 deaths, the relative risk of A/B may be as high as 20:1 or as low as 1:5 — assessing both risks "conservatively" or via "best estimates" does not alter the fact that either could be larger than the other).

Together, uncertainty and variability are arguably much more serious than the "apple and orange" problem. If you know your choice is between an apple and an orange, you can probably make a reasonable choice even though you will get more of some desired dimensions and less of others. Yet, if you can't measure or even reliably estimate, e.g., the price, calories or taste of the fruit, you can't choose between even two apples with any assurance you will pick the "better" one.

Fortunately, when comparing risks the influence of uncertainty and variability on the comparison can often be roughly quantified with only

---

25 *See* Appendix A.

a marginal increase in the analytic effort needed to generate point estimates of risk. *Then, comparing risks thoughtfully simply requires rephrasing the questions asked.* For uncertain risks, the question "Which risk is larger?" needs to be refined into "How confident can we be as to which risk is larger, by how much (with what consequences if I guess wrong)?" Appendix A summarizes procedures for asking this latter question and shows some of the different decision problems that result. For risks that vary spatially, temporally or inter-individually, the question needs to be rephrased as "for *whom* (or when or where) is this risk larger?" In the "groundlings" example discussed below, the answer to this latter question might be "pesticide exposure is riskier than being out-of-doors where an airplane might crash on you, unless you live within X miles of an airport flight path." Similarly, Schwing and Kamerud showed that the question "Is it safer to drive or to fly?" depends in large part on the day of the week and the time of day you plan to travel, as the individual fatality risks in automobiles vary across the 168 hours in each week by a factor of 134.[26]

Basic Theorems of an Improved CRA

The raw materials for broadening the scope (hence public acceptability) of CRA consist of a fuller enumeration of risk attributes. The next section will present such a listing of "dimensions" by which risks differ, including various ways to express the statistical magnitude of risks as well as many other dimensions unrelated to magnitude. However, a "laundry list" alone would be of little incremental value without two other improvements: A conceptual framework for applying the "multi-dimensional" mindset to actual risk comparisons, and a social and institutional process by which such a framework could be implemented. These topics will be treated in turn, as "bookends" on either side of listing the dimensions themselves.

To move beyond the "ten deaths are more than nine deaths" kind of one-dimensional comparisons that characterize the current state of CRA, I propose three related theorems:

---

[26] Richard C. Schwing & Dana B. Kamerud, *The Distribution of Risks: Vehicle Occupant Fatalities and Time of the Week,* 8 Risk Anal. 127 (1988).

- *Theorem 1:* Choices (the more general type of decision problem) or risks (the type of problem relevant here) ideally should be compared along all of the relevant dimensions that differentiate them.
- *Theorem 2:* If some of these dimensions (in the limit, all but one of them) are to be ignored for the sake of simplicity, time or other constraints, the remaining dimension(s) should be the most *important* one(s).
- *Theorem 3:* The *importance* of a dimension is in turn a function of two factors: The absolute amount by which the choices or risks differ with respect to this dimension (an empirical question), and the relative influence of the dimension on the decision (a value judgment).

To illustrate how these ideas apply in practice, let us consider two stereotypical comparisons: The literal choice between an apple and an orange, and the comparison between the risk of being killed by a crashing airplane (while you are on the ground) and the risk of getting cancer from pesticide residues on your food. In the apple/orange comparison, if resources did not permit making the comparison along all of the potentially relevant dimensions listed at the beginning of Table 2, then Theorem 3 provides guidance as to which of many possible simple comparisons are apt to be uninformative or misleading.

The first part of Theorem 3 cautions the analyst against focusing the comparison on a single dimension where the absolute difference between the two risks may be either *statistically insignificant* (i.e., the "signal" that one risk surpasses the other along this dimension is smaller than the "noise" that makes us unable to measure either risk precisely) or *functionally insignificant* (i.e., the difference may be real and definitive, but it is so small as to be of no practical importance).[27] An example of this kind of "tunnel vision" would be comparing apples and oranges based solely on the assertion that apples were lower in calories and therefore preferable, if in fact the average apple had 69 calories to the orange's 70 (and even more so if inaccuracies in measurement or the variation in calories among apples could outstrip this one-calorie across-species average difference). The second half of Theorem 3 offers a

---

27   In The Prelude, bk. II, l. 216-219, William Wordsworth was particularly critical of over-reliance on
>   that false secondary power/By which we multiply distinctions, then/Deem that our puny boundaries are things/That we perceive, and not that we have made.

different caution: Even if a difference is definitive and substantial, one ought not to highlight a distinction that does not or should not strongly influence the overall comparison. Suppose, for example, that apples were markedly enriched in Vitamin K relative to oranges. It would then be accurate, but manipulative, to advise someone to buy apples solely on the basis of this real but probably irrelevant difference.

These two pitfalls seem obvious in the context of buying fruit, but they are routinely ignored in comparing environmental, health and safety risks. Consider the second example — that of the airline risk to "groundlings" versus the cancer risk from some particular pesticide. It is certainly possible to argue that in a given case, the former risk is worse than the latter by virtue of a larger probability of fatality. John Graham has testified to this effect before Congress on several occasions,[28] citing the empirical work of Goldstein et al.[29] According to these experts, the average U.S. citizen faces a lifetime risk of about 5×10^{-6} of being killed by being hit by a crashing airplane. Thus, a 1×10^{-6} lifetime excess risk standard applied to carcinogenic pesticides is arguably irrational because it seeks to reduce risk to a level one fifth as large as one perceived as remote and already "accepted" by society (whose citizens seem to be unconcerned about being out of doors or in unreinforced structures where an airplane might fall on them).

However, such a comparison conceivably violates both of the precepts in Theorem 3. First, the difference between five per million and one per million may be a "puny boundary" á la Wordsworth: We are used to thinking of a factor of five as a significant difference, but numbers on the order of 10^{-6} are not nearly so precise or directly verifiable as the kinds of quantities that reinforce the belief that a five-fold difference is large. More to the point, these average risks do not reflect the substantial variation in risk that real people undergo; the reason that most people probably think the risk of being hit by a falling airplane is remotely low is that for most of us it *is* remotely low. That

---

[28] John D. Graham, testimony before joint hearing of the House Subcomm. on Health and Environment of the Comm. on Energy and Commerce and the Senate Comm. on Labor and Human Resources, Sept. 21, 1993; also *The Role of Risk Analysis in Environmental Protection,* his testimony before the House Comm. on Government Operations, Feb. 1, 1994.

[29] Bernard D. Goldstein et al., *Risk to Groundlings of Death Due to Airplane Accidents: A Risk Communication Tool,* 12 Risk Anal. 339 (1992).

average of 5×10^{-6} is probably a hybrid of two risks: A higher risk among the small minority of people who live near airport flight paths (and especially near airports, since the risk of a crash is greatest at takeoff and landing), and a much lower risk among the majority of us.[30] For most U.S. citizens, therefore, the one per million risk from the hypothetical pesticide may in fact be *larger* than the risk of the airplane crashing on them.

Even if the numerical distinction between the two types of risk were definitive and applied across-the-board, the second pitfall in Theorem 3 might apply — that the numerical dimension was unimportant compared to other unexamined dimensions. Being hit by an airplane and contracting cancer from pesticides are both involuntary; the ultimate consequences might be similar, but the risks differ in many ways. The next section enumerates many of these; for now, consider the single extra dimension of offsetting benefits. People may logically regard having a national system of frequent airline flights as a benefit making annual several deaths among "groundlings" acceptable, but they may also feel that pesticide benefits do not justify a commensurate or even a smaller number of excess deaths. I agree with Graham's testimony that we should carefully consider the absolute benefits of a pesticide before we vow to reduce or eliminate what may be its small risks. Doing so, however, means that the relative comparison between the *benefits* of airplane travel and pesticides may push the numerical comparison of statistical risks to a subsidiary or even an irrelevant position. In the language of Theorem 3, it may well be that the dimension of concomitant benefits is both more *significant* and more *influential*, i.e., more "important" than turning on the "lamppost" to illuminate fatalities alone.

Enumerating the Dimensions of Risk

On the one hand, any list of the potentially important dimensions by which risks may differ should be taken with a grain of salt, as it necessarily reflects strongly the subjective personal views of the author.

---

[30]   For example, if 98% of the deaths to "groundlings" occur among the one million people who live nearest to airports, their risk would be 1.2×10^{-3}/lifetime, but the risk to the remaining 249 million U.S. citizens would be 10^{-7}, one tenth of the hypothetical pesticide risk.

On the other hand, it is unfortunate that few such lists have been published to date,[31] because only through comparing the idiosyncratic views of different observers can we begin to synthesize and codify a list of dimensions that citizens and policymakers could use in common.

In preparing the list that follows, I was guided by rules proposed by Morgan et al.[32] First, they advised that the dimensions (what they call "attributes") listed should comprise a set that is: (1) comprehensive, (2) non-redundant, (3) preferentially independent, i.e., a person's judgment about the amount of a particular dimension for a given risk should not depend upon amounts along other dimensions, (4) measurable and (5) minimal in number, i.e., not needlessly detailed. Moreover, they recommended that the attributes should be sorted into a small number of categories and that such sorting should be guided by a principle of correlation, i.e., for any given risk, people's judgments about various attributes that fall *within* the same category should be fairly well-correlated, whereas their judgments about other attributes that cut *across* categories should be rather uncorrelated.[33]

I have sorted my own conception of a "minimal set" of dimensions into three categories: "magnitude," "dread" and "social context." Two of these are very similar to the three groups of attributes Morgan et al. developed.[34] However, within each category I list various dimensions that do not appear in that typology — and vice versa.

Category I: Magnitude

•   *Unweighted population-based measures of magnitude.* An example is the number of excess fatalities (or cases of disease or injury) per annum or per lifetime. This is by far the most prevalent way of

[31] Although others have written about the "multidimensionality" of risk, they have tended only to present highly aggregated combinations of dimensions, not the kind of major and minor typology presented here. For example, the well-known "risk mapping" pioneered by Slovic highlights only two dimensions of risk — albeit very inclusive ones — "degree to which the risk is known" and "dread." *See* Paul Slovic, *Perception of Risk,* 236 Science 280 (1987).

[32] M. Granger Morgan et al., *A Procedure for Ranking Risk within Federal Agencies,* in Comparing Environmental Risks: Tools for Setting Governmental Priorities (J. Clarence Davies, ed. 1995), at 111.

[33] For example, in the literal apple/orange comparison modeled in Table 2, there might be a category of aesthetic attributes that were closely correlated with each other, but another category of economic attributes that each were not correlated with any of the aesthetic judgments.

[34] M. Granger Morgan et al., *supra* note 32.

comparing risk magnitudes; it is so far the only measure EPA uses to inform its "large" risk comparisons.

• *Weighted population-based measures that do not treat all fatalities as equivalent.* An example is the commonly-used "person-years of life lost," which weights the death of an infant much more heavily than the death of an octogenarian.

• *Individual-risk measures independent of the number of persons at risk.* These measures are in units of probability, not of consequence (e.g., the probability that an individual with average exposure will contract cancer over a year or a lifetime). Generally, EPA and others try to estimate the risk to a real or hypothetical individual whose exposure (although not susceptibility) is above-average or "maximal". To oversimplify, the population-based measures imply a "public health" philosophy of risk management — that the worst risks affect the greatest number of persons — while individual-risk measures imply a "civil rights" philosophy — that the worst risks are those that subject any persons to the highest probabilities of death or disease.[35]

• *Hybrid measures that incorporate characteristics of both population and individual risk criteria.* Estimating the magnitude of risk either by population-wide consequence alone or by maximum (or average) individual risk alone necessarily discriminates against people who live in either densely-populated or sparsely-populated areas. The former type of measure regards a risk of 10^{-6}/person in a city of two million as more serious than one of 10^{-2}/person in a village of 100. So the villagers might receive less protection even though their risk levels was each 10,000 times higher than the urbanites' (i.e., two expected fatalities in the city versus one in the village). On the other hand, if the individual-risk criterion were used exclusively, urbanites would lose out in the sense that their risks might have been reduced twice as efficiently on a cost-per-life-saved basis than could the villagers' risks. Yet, such discrimination based on population density does not have to occur. As several others have pointed out,[36] even a very simple hybrid measure

---

[35] Adam M. Finkel, *A Way Out of the "Individuals versus Populations" Dilemma in Air Toxics Regulation,* unpublished, Center for Risk Management, Resources for the Future (1990).

[36] Finkel *supra* and Paul Milvy, *A General Guideline for Management of Risk from Carcinogens,* 6 Risk Anal. 69 (1986).

.

such as the product of expected fatalities (D) and maximum individual risk (R_{max}) can help balance competing social goals. For example, a goal of reducing (D x R_{max}) to below 0.0001 would imply, each on the borderline of "unacceptability," equivalence of: 100 persons, each facing a lifetime risk of 10^{-3} (D = 0.1, R_{max} = 0.001) and one million, each facing a risk of 10^{-5} (D = 10, R_{max} = 0.00001).[37] By weighting or otherwise modifying this product, risk managers can strike any other balance between desire for efficiency and aversion to inequality.

• *Measures that incorporate the concept of "background."* Some experts believe that society's evaluation of a particular excess risk should not be made independently of pre-existing (and perhaps unavoidable) sources of the same risk.[38] By this reckoning, an incremental risk of (say) 10^{-5}/lifetime caused by disposal of uranium mill tailings would be a higher priority in a community where there were no other sources of exposure to uranium decay products than in an area where the natural background of radioactivity in the soil was already subjecting residents to risks on the order of 10^{-4} or higher. In addition to any of the measures discussed above, therefore, risks could be compared based on their relative or absolute increment above the background.

Category II: Dread

• *Degree of fear.* This is one of the most significant dimensions of risk, and it varies widely both across individuals within a society and cross-culturally. Fear can help define what the most important risks are risks *of.* For example, U.S. citizens tend to regard a lingering death as more dreadful than a sudden demise. Yet, the opposite is apparently true in some Asian nations, where the "worst" death is a sudden one that occurs away from home where one's family and ancestral spirits cannot help ensure safe passage to the afterlife. Fear can also be an adverse outcome per se, thereby adding to the graveness of certain risks that are highly feared.

---

[37] Note that the first risk would rank as 100 times more dire than the second if only the maximum individual risks were compared, while it would rank one-hundredth as serious if deaths alone were compared. If the reader believes that the two situations in fact seem to be of roughly equal severity, then an existing ranking procedure that could produce two answers that differ in sign and by 10,000-fold in magnitude ought to seem quite suspect.

[38] Frank B. Cross, Daniel M. Byrd III & Lester B. Lave, *Discernible Risk — A Proposed Standard for Significant Risk in Carcinogen Regulation,* 43 Admin. L.Rev. 61 (1991).

• *Degree of irreversibility.* Irreversibility generally increases the perceived severity of a risk, whether the continuum from rapid reversibility to complete irreversibility applies to injury (even major lacerations are generally seen as less severe than spinal cord injuries), disease (gastroenteritis versus cancer) or ecological harm (an oil spill versus a species extinction).

• *Degree of individual controllability.* Two risks of equal magnitude (however that is measured) may elicit strikingly different individual or social responses if one risk can be avoided or reduced through personal action and the other cannot (readily or at all). For example, basketball may be perceived as more hazardous (if injury rates were comparable) than football, because children cannot realistically compete in the former sport wearing substantial protective equipment. Similarly, a carcinogenic contaminant would probably cause more concern if it was present in milk than in maraschino cherries.

• *Degree of deferral to future generations.* Individuals may be strongly affected by the time course of adverse effects. For some, risks that manifest themselves only in subsequent generations are particularly dire (because of possible repugnance for making posterity pay); for others, delayed risks are much less important (because of "discounting" future costs and the possibility of technological remedies available by the time effects are manifested) — another striking example of an attribute (time) being either directly or inversely related to benefit, depending on individual perception.

Category III: Social Context

• *Salience of blame.* People desire that risks be reduced for various reasons, some that transcend the direct and tangible benefits to themselves (e.g., increased life span). One of the most influential of these intangible factors, in my opinion, is the desire to focus attention on reducing risks where in so doing injustices can also be redressed. Scholars have advanced numerous explanations, for instance, why people generally seem more concerned about the ostensibly smaller risks from proximity to a hazardous waste disposal site than they are about the larger risks from radon gas in their homes. While a small part of the disparity may be explained by the fact that radon is colorless and odorless, and a larger part by the reality that Superfund cleanups are at

least perceived to come out of someone else's finances, I suspect that the most important factor of all is the sense that someone is to blame for the exposure to waste. The positive action (to dump rather than reuse or treat) sets in motion the desire for an affirmative counter-action, a calculus that does not apply in the case of radon.

• *Degree of identifiability of those at risk.* Much has been written about our societal propensity to devote much more than proportionate attention and resources to reducing risks to identifiable victims, be they three astronauts stuck in the malfunctioning Apollo 13 spacecraft or three whales trapped in the Alaskan ice pack. A risk that affects a few readily-identifiable victims (such as some chemical that might aggravate respiratory symptoms in asthmatics) thus might rate more concern than another chemical that affects a larger number of persons seemingly at random from out of the entire population.

• *Benefits of the risky activity or exposure.* As described in the previous section, using the example of being hit by a falling airplane versus contracting cancer from exposure to a pesticide, looking at risk is at most only half the story. Some risks are more worth taking — or bearing — than others, and this difference is largely governed by the perceived benefits that accompany the risk.

• *Cost and feasibility of reducing risk.* Thus, added to the question "How large is the risk?" is the concomitant question "Why should we not wish to reduce it?" The answer depends crucially on what benefits would be abrogated if the risk was reduced or eliminated. The equation cannot be complete, however, until we accompany *"Should* we reduce?" with *"Can* we reduce?" Cheaply and easily eliminated, small risks may deserve priority, whereas in a thorough risk comparison, even very large risks may not — if reducing them would be technically infeasible or prohibitively expensive. This dimension is often crucial because different risks can vary enormously when it comes time to compare interventions as well as hazards. For example, our perceived ranking of the need to control handguns in schools would differ dramatically, I contend, depending on whether metal detectors cost $100 or $1 million, or if the technology did not exist at all.

• *Risks of the intervention itself.* The two dimensions discussed immediately above form a nearly-complete accounting of the basic

reasons for tolerating any risk — if it provides benefits as well as risks, and if it would be difficult or impossible to eliminate it. There is one final reason which raises issues of both benefits and cost/feasibility — the dimension of "offsetting risks," which arises whenever reducing one risk would create new risks. One can therefore look at any "primary risk" in a new light — that part of its accompanying benefit is freeing us from some other risk, or alternatively, that its elimination might be costly in units of risk as well as dollars. There are many ways in which risk reductions can themselves be risky. Graham and Wiener develop a typology that categorizes both the type of risk created (similar to or different from the primary risk) and the population affected (the same as or different from those affected by the primary risk).[39] For example, expelling violent students might transfer a similar risk from one population (in-school victims) to another (members of the local community). Or, eliminating the chlorination of water to reduce possible cancer risks in schools might create a new risk (from pathogens that survive less effective means of disinfection) among the same population. In any case, the implicit ranking given to the primary risk might well be strongly influenced by the "risk consequences" of the feasible ways of addressing it.

Social and Informational Processes for Comparing Risks

We compare risks — at least, the reason we ought to — to help guide our actions. We need a way to implement the three theorems presented above (or some superior conceptual framework for comparing risks) so that the affected stakeholders can reach a consensual determination about which risks are most worthy of attention. Whatever process society chooses for putting comparative risk assessment into practice, it ought to advance two distinct goals in turn.

> (1) *Provide a forum for identifying, and making judgments about, the "important" dimensions of the risks being compared.* As a practical matter, the process should therefore move the debate over which risks are "worst" beyond the current "tunnel vision" that only considers point estimates of consequences, and yet not lurch so far towards an ornate characterization of each risk that the debate

---

[39] Risk Versus Risk: Tradeoffs in Protecting Health and the Environment (John D. Graham & Jonathan B. Wiener, eds. 1995).

becomes unmanageably complex and divorced from any consideration of quantitative risk information.

(2) *Provide a framework for asking, and moving towards consensus about, the real underlying question: "What should we do to make our lives safer and do less damage to the environment, given that any intervention we undertake will use up resources from a finite supply?"* This is the question that elevates CRA from a fascinating but ultimately futile endeavor (of deciding what is most worth worrying about) to a complete and practical activity of deciding what we should do, given what we can do and what it costs to do it.

The remainder of this article will discuss two related process issues: How to structure the public debate over setting priorities to reduce risks; and what information analysts should transmit to the involved parties to initiate or inform such debate.

Considerations of Social Process

Most of my insights into what kind of process might be optimal for coming to public judgment about the relative risks of various hazards are drawn from a 1992 Center for Risk Management conference entitled "Setting National Environmental Priorities: The EPA Risk-Based Paradigm and its Alternatives."[40] It dealt exclusively with "large" CRA, and even more specifically with the intent of then-Administrator William Reilly to reorient EPA's budget with the goal of "reducing the worst risks first." However, the discussions about the possible processes for judging what to do first are, I think, broadly applicable to "small" CRA as well.[41]

Much discussion at the conference centered around the distinctions between the so-called "hard" version of risk-based priority setting championed by the U.S. Office of Management and Budget (and, according to some, EPA as well) and the "soft" version preferred by some other stakeholders. Many believe that the "hard" version — in which a small group of experts estimate the magnitude of various risks and marginal costs of reduction to yield a numerical ranking of risk

---

[40] Worst Things First? The Debate over Risk-Based National Environmental Priorities (Adam M. Finkel & Dominic Golding, eds. 1994).

[41] The 1992 conference was not intended to generate consensus among the 100-plus participants, so any of the opinions paraphrased here are merely ideas that some of the attendees put forth or endorsed.

reduction opportunities in order of "bang for the buck" — can do more harm than good. Certainly, this article suggests that confining the ranking process to "experts," and further circumscribing it to deal only in the currency of "risk numbers," will not be productive in advancing social judgments on risks for two overriding reasons: (1) The conceptual ranking tool used is one-dimensional when many other dimensions may be of equal or greater "importance" than risk magnitude alone; (2) even if magnitude is the most important dimension, the dynamic of monologue rather than dialogue to determine the ranking will tend to foment resentment and mistrust among the affected citizens.

The "soft" version is no panacea, however. In this paradigm, a representative group composed of citizens and "experts" would work together to generate a more "impressionistic" ranking out of a consensual weighting of the various dimensions that distinguish the risks under consideration. The obvious objection to the "softening" of CRA is that it allows people to make "soft" dimensions as important as, or more important than, the quantitative information about how to reduce risks most significantly with the resources we have — in essence, the "soft" version may just be a polite way to describe the overly emotional, haphazard, inefficient way we now set priorities. A perhaps less obvious, but potentially more damaging criticism was raised by Donald Hornstein at the 1992 conference.[42] Hornstein and others have pointed out an irony — that the "soft" version is held up as an alternative to the technocratic elitism of the "hard" approach. Yet, it may be no less vulnerable to being dominated by special interests. A system that purports to examine risks holistically will fail to live up to that noble aim if in practice the appointed ranking group (or a powerful subset thereof) successfully redefines risk as "fear" or "inequity" or "outrage" or some other dimension that may be just as sterile in its own way as "magnitude" alone.

Thus, besides the challenges facing us to improve the way we characterize risks, we have yet to put forward a means of aggregating and reconciling the unique ways each citizen will choose to process that information to arrive at a comparative risk ranking. Those who wish to

---

[42] *Paradigms, Process, and Politics: Risk and Regulatory Design,* in Worst Things First? *supra* note 40, at 152.

advance the debate over whether and how we should change the way we manage risks need to be mindful that better information alone will not improve whatever suboptimal priorities we have — without a better mechanism for translating that information into conclusions all parties can accept (or at least can live with).

Considerations of Information

Regardless of whether comparative risk information eventually is transmitted to "experts," a group of laypeople intended to represent the broader public, or even the whole citizenry (perhaps as a prelude to a referendum), I contend that the subsequent process is more likely to yield both reasonable and publicly-acceptable outputs if four precepts regarding the type of information proffered are considered.

(1) *Tailor the level of detail (e.g., the number of dimensions discussed) to the purpose of the comparison: More important comparisons deserve more detailed information.* The concept of "iteration" — moving beyond a "one size fits all" approach to risk assessment by matching the ambitiousness of the assessment to the needs of the consumers — may be taking hold as a principle for the future of the field.[43] Iteration applied to risk comparisons would mean that when the decision is straightforward enough or sufficiently routine that a "back of the envelope" characterization will suffice, an exhaustive and ornate comparison would be wasteful. Conversely, for complex and socially important decisions, the richness of the comparison should reflect its intended use. A potential example of a risk comparison where a less detailed set of information might well suffice is that of whether to equip school buses with seat belts; here, the influence of dread and social-context attributes may be minor enough that a "harder" quantification might tell most of the story. On the other hand, deciding whether to embark on either an ambitious program of asbestos mitigation in schools or to devote the same resources to violence reduction programs in schools is such a weighty and value-laden choice that the more dimensions analyzed and discussed with the stakeholders, the better. Other things being equal, I believe it is also important to follow the principle that the more the authors of a comparison intend it to *convince* rather than merely inform recipients,

[43] *See* Science and Judgment in Risk Assessment, *supra* note 23.

the more detailed and multi-dimensional the comparison should be. It should require less care to inform people that they can rationally judge either of two risks to be more dire than to lead them towards a conclusion, however well-grounded, that one or another risk is worse.

(2) *Express risks both as population consequences and as individual probabilities; highlight the uncertainty in the former and the variability in the latter.* There is still a vigorous debate about how to fully characterize uncertainty and variability in any given risk assessment.[44] Also, of course, even knowing how to do it exhaustively would not solve the practical problem of how to manage the tension between completeness and comprehensibility. As we move towards a better understanding of the twin phenomena of uncertainty and variability, I would suggest a rough schema for reporting risk estimates (the "magnitude" dimension of risk comparisons). For "detailed" comparisons (see the discussion of "iteration" above), this scheme would replace the single number currently used (Risk A is X times as big as Risk B) with a manageable alternative set of six numbers:

- The lower and upper bounds[45] on R, the ratio of one population-wide risk to the other. So, comparison of the consequences of two risks would take the form "Risk A causes between ten and 100 times more fatalities annually than Risk B" or "with 95% confidence, all we can say is that A causes somewhere between ten times more than and half as many fatalities as B" (see Appendix A).
- The lower and upper bounds on the individual risk posed by Risk A and by Risk B (i.e., four more numbers, for a total of six), accounting for inter-individual variations in such risk. At a minimum, this range will allow individuals with differing attitudes towards risk to see the full panoply of possibilities they might face and judge their prospects accordingly. Much more helpful if the information is obtainable, though, would be to supplement the bounds with information on what distinguishing characteristics of individuals correspond to each of the bounds. For example, the probability of being a victim of some type of violence might range from 10^{-6} (in some particular rural community)

---

44 *Id.*

45 These bounds should probably be the statisticians' standard 5th and 95th percentiles, or a slightly more inclusive or less inclusive pair (e.g., the 10th and 90th), rather than more extreme combinations such as the quartiles (25th and 75th) or the 1st and 99th.

to 10^{-3} (in the center of some troubled urban area). Giving both the numbers and their geographical correlates would be far more informative than simply communicating the numbers above. Or (to give an example with a continuous correlate), the cancer risk from some pesticide in food might also range from 10^{-6} to 10^{-3} because the intrinsic risk was 10^{-6} per gram of the food consumed per day (and because 5% of the population eats one kilogram of the foodstuff per day). Both examples illustrate what I believe is an absolutely critical concept in risk communication: Citizens care about the risks to themselves or to other "real people," not a hypothetical average or worst-case person. *If analysts can help people narrow down the risk range to a meaningful estimate specific to their unique circumstances, then they will have truly engaged in "risk communication."*

(3) *Compare actions, not disembodied risks. Setting priorities is more than simply ranking risks.* As many have remarked,[46] to set priorities means to guide where *resources* should flow; while the "biggest" problems may be mental priorities, they may bear no resemblance to functional priorities. Large risks may have no feasible, economical or politically acceptable means of control or prevention, while small risks may be eliminated through actions that carry a small or even a negative economic price. Therefore, even if none of the psychosocial and contextual dimensions of risk are to be included in the analyst's attempt at risk comparison, decision-makers and stakeholders need information on the costs and feasibilities of *specific interventions* to judge where resources should flow. These estimates may be as uncertain as the risk estimates are, and may add further complexity to the social process, but the alternative is either to rank the risks alone and have no guide for policy, or (perhaps worse) for decision makers to assume that the risk ranking equals the resource allocation.

(4) *Regard the initial information that flows from the analyst to the decision-maker or to the populace as just that — initial.* In addition to having a central role to play in evaluating the empirical and narrative information about the various dimensions of the risks being compared, the stakeholders also may have much to contribute in structuring and supplementing the information itself. Depending on

[46] *See, e.g.,* Dale Hattis & Robert L. Goble, *Current Priotity-Setting Methodology: Too Little Rationality or Too Much?* in Worst Things First? *supra* note 40.

the circumstances, the "consumers" of comparative risk information may wish to see other dimensions analyzed — or may even want to consider other risks (or to aggregate the given set of risks into different categories). Their most far-reaching role might well be to impel the comparison of risk reduction interventions not initially considered. For example, EPA's "large" risk comparison effort has been criticized for only trying to move society "out of the fire and into the frying pan."[47] Hornstein claims that in comparing, among many other examples, the cost-effectiveness of reducing air pollution risks by shifting from fossil fuels to nuclear power (versus the increase in safety hazards and other risks that the shift might achieve), EPA has discouraged society from asking the larger question: Can we reduce demand for both sources of electricity simultaneously, or shift to a power source that is superior in terms of both emissions and safety?

Similarly, some criticize "zero-sum" choices inherent in government's declaring it can address only one risk or another.[48] While current agency budgets might be sufficiently tight that expenditures on one risk must crowd out efforts on another, the "first-best" solution might well be to redefine the boundaries of the problem and see if other governmental or societal expenditures are even less efficient than either of the two competing interventions. In other words, one response to the assertion that we need to spend less on Superfund cleanups to furnish smoke detectors to all public housing units (a shift that would almost certainty save more "statistical lives") is that there are many other "pockets" from which the smaller expenditure for smoke detectors could be drawn. It is very important that in many cases a more thorough examination of such purported "win/win situations" might reveal that there is in fact "no free lunch." Yet, I expect that the question needs to be asked more frequently and that stakeholders will view it with greater trust if risks are not automatically played off against one another. At the end, we compare risks to reduce them — not to fool ourselves that such reductions are painless or to cringe before the opportunities to make our lives safer and healthier.

---

47   Donald T. Hornstein, *Reclaiming Environmental Law: A Normative Critique of Comparative Risk Analysis*, 92 Colum. L.Rev. 562, 626 (1992).

48   *See, e.g.,* Mary O'Brien, *A Proposal to Address, Rather Than Rank Environmental Problems,* in Worst Things First? *supra* note 40.

Appendix A
The Mathematics and Logic of Comparing Uncertain Risks

Uncertain quantities, including individual or population risk estimates, can be summarized via any of several "estimators" or "summary statistics." The most well-known of these are the median (the 50th percentile), the mean (also known as the expected value), and some upper confidence limit such as the 95th or 99th percentile. These summary statistics are very important in compactly conveying information in a form readily accessible to decision-makers and the public. When comparing two uncertain risks, however, any of these summaries can give a misleading picture of the relative size of the risks; in fact, no estimator alone is sufficient even to determine qualitatively which risk is larger and which is smaller. Moreover, the highly-touted "best estimates" (usually either the median or mean) are no better than any other single measure at avoiding highly misleading results, despite connotations of the adjective "best" and the tendency of many CRA practitioners to use these measures to make facile comparisons while criticizing all other measures as "value-laden." The following example shows how misleading summary statistics can be, and suggests ways to communicate the complexity inherent in correctly comparing risks.

Consider two substances which can both cause cancer in humans; assume further that in a nation of ten million inhabitants, each person is exposed to the same amount of each substance. The only uncertainty in individual or population risk is thus due to our lack of knowledge of the exact potency of either substance.[49]

Suppose our "best estimate" of the potency of Substance A is such that it would pose an excess individual risk of 5×10^{-5} (i.e., one excess death per 20,000 persons exposed over a lifetime, or 500 excess deaths in the hypothetical country every 70 years). However, suppose scientists are sufficiently uncertain about A's potency that they believe there is a 5% chance it is about 5.2 times more potent than its "best estimate" and a corresponding 5% chance it is 1/5.2 times as potent.[50]

[49] For brevity, I present only this one example. The mathematics would be unchanged if I replaced all the sources of uncertainty with sources of inter-individual variability, although the social implications of the results might not be. I encourage the reader to re-read this example as if all the imprecision was due to differences among individuals rather than to scientific uncertainty.

[50] This "uncertainty factor" of 5.2 is arguably quite small compared to many

Now suppose our "best estimate" of the potency of Substance B is that it is *one fifth as large as* that of A; in other words, we believe the individual risk from B is 1 x 10^{-5} (or 100 excess deaths every 70 years). We know a bit less about B than we know about A; it could be 29 times more than or 1/29th as potent as we think (again, each alternative has a 5% chance of being correct).

Let us look at four different ways to compare the two risks, each based on a different summary measure, to see how difficult it can be to choose the "right" way to compare them:

(1) Comparing the two "best estimates" leads us to believe that A is five times riskier than B, as seen above (500 versus 100 deaths).

(2) But consider another type of "best estimate." Rather than the median used above, we could use the mean, which is the sum of all possible values of risk weighed by their probabilities.[51] I have constructed this example such that, judged via their means, the two risks are exactly equal (both 8.24 x 10^{-5} on an individual-risk basis, or 824 excess deaths expected in the whole population). Here, each mean value is higher than its median because of the fact that the chance we've underestimated the risk by X-fold has more of an effect on the estimate than the equal chance we've overestimated it by X-fold — just as in the note below, where the people who earn $500,000 are "richer than average" by a larger amount than people who earn $5,000 are "poorer than average."

(3) Suppose you were particularly concerned about the "reasonable worst-case" values of each risk. Here B is more uncertain than A, and in fact B is riskier than A by the criterion of the respective 95th percentile estimates (there is a 5% chance the risk of A is 2.6 x 10^{-4} or greater, while for B the corresponding estimate is 2.9 x 10^{-4}).

---

imprecisions that characterize actual cancer risk assessments. For purists, this factor is actually a logarithmic standard deviation of 1.0 from an underlying lognormal distribution whose median equals the "best estimate."

[51] The median is an average that doesn't account for how far away from the center any other value is; each measurement (or person) counts equally. In a group where ten people have annual incomes of $5,000, 80 earn $50,000, and ten earn $500,000, the median is $50,000 because as many people earn this amount or more as earn less. The mean, however, would be [10(5000) + 80(50,000) + 10(500,000)] ÷ 100, or $90,500. Both measures are informative in a different way. The former is akin to looking for an "average person" and then asking her what she earns; the latter is more of a best guess of what a random person would earn, or what the population on average earns. In a lognormal distribution, the mean exceeds the median by a multiplicative factor of $\exp(0.5\sigma^2)$, where σ is the logarithmic standard deviation.

(4) By the 95th percentile criterion, B is admittedly not substantially worse than A. But why consider only the 95th percentile? Suppose you wanted to know how high each risk might be and still have a one in 100 chance of being even higher. By this 99th percentile criterion, B (whose 99th percentile estimate is 1.2×10^{-3}) is more than twice as risky as A (which is equally unlikely to be higher than 5.2×10^{-4}). The more concerned you were about incorrectly underestimating either risk, the riskier B looks.

Consider these four comparisons in tabular form:

Estimator	Risk A	Risk B	Ratio A/B
Median	5×10^{-5}	1×10^{-5}	5:1
Mean	8.24×10^{-5}	8.24×10^{-5}	1:1
95th percentile	2.6×10^{-4}	2.9×10^{-4}	(1:1.1)
99th percentile	5.2×10^{-4}	1.2×10^{-3}	(1:2.3)

As this indicates, depending on which summary statistic you use, the two risks can appear to be equal, in one rank order, or in the opposite rank order.

Since we cannot determine confidently which risk is larger, how can this ambiguity be conveyed informatively? After all, there is a substantial difference between "we cannot tell which risk is larger" and "we know that the two risks are equal." The former situation applies here, not the latter.

One promising way to attack this problem is to focus not on the uncertainty in the absolute magnitude of either risk, but on the uncertainty in the *ratio of one risk divided by the other*. In this hypothetical case, the quantity (B ÷ A) is distributed normally (a "bell-shaped curve") on a multiplicative scale. To be precise, the center of this distribution lies at the point 0.2 (that is, A is five times larger than B, as the two median estimates in the table indicate), but the 5th and 95th percentiles of the ratio are a factor of 42.8 away from the center. In other words, A could be 214 (5 x 42.8) times greater than B, but with equal probability B could be 8.6 (1/5 x 42.8) times greater than A.[52]

[52] Note that the 95th percentile of the ratio (B/A = 8.6) does not equal the value in the third column of the table (B/A = 1.1), which is the quotient of the two 95th percentile values of each risk viewed separately. This again shows how only comparing separate summary statistics can give misleading results.

One can also examine this ratio uncertainty to find the point on the distribution where the two risks are equal; in this case, there is about a 75% chance that A > B, not an unexpected result since at the center of the distribution A is five times larger than B.

Now we come to a crossroads. The single number as an answer can lead one headlong into a wrong decision. Yet, compared to the more cumbersome analysis and communication of the uncertainty in the ratio, at least it is clear and comprehensible. While the subject of "rational" decision-making under uncertainty is a complex and controversial one, let me suggest three paradigm situations where the more thorough kind of comparison advocated here can be both efficient and informative. Together, these situations cover most if not all of the real cases decision-makers and citizens will encounter when risks must be compared.

Case 1: *One risk is clearly larger than the other over most or all of the ratio distribution.* This situation will occur either when the central estimates diverge wildly or when the uncertainties are very small (or both, of course). Here the "right way" to do the comparison yields the same answer as the facile way, but the former is more informative and more credible or trust-engendering. To say "Risk C is five times larger than Risk D" is only slightly more compact than to say "we are 99% sure Risk C is between three and ten times larger than Risk D," and the latter statement avoids conveying a false sense of precision.

Case 2: *The "noise" in the comparison (far) outweighs any "signal" of relative risk.* This paradigm situation is more common than many "experts" in CRA realize. For example, I have argued[53] that the oft-quoted result that "aflatoxin is eighteen times riskier than Alar" was doubly misleading because the central estimate of the ratio was much closer to 1:1, and, much more importantly, because the uncertainty spanned four orders of magnitude, such that either risk could well have been ten to 100 times larger than the other. Here the superiority of the more thorough approach to CRA is most obvious. To say "Either risk might be much larger than the other, but we can't tell which one" is not at all an admission of defeat — it is equivalent to saying "We've looked for the keys under the lamppost and haven't found them there."

---

53 Finkel, *supra* note 22.

It puts a stop to a fruitless search and impels us to look elsewhere. That "elsewhere" may be to improve the data on exposure or the scientific basis of the toxicity information, so that in the future we *can* pick the "signal" out of the "noise." Or, it may be to declare that we cannot resolve the comparison via the single dimension of statistical magnitude and must therefore look to other ways in which the two risks may well differ more meaningfully and definitively.

Case 3: *There is both "signal" and "noise" to reckon with.* Here is where the real discipline and craft of risk management must come to the fore. When one risk is not unambiguously larger than the other, and yet a choice must be made, any choice may turn out incorrect. Risk management is about balancing the probabilities and consequences of such errors. For example, in the hypothetical above, how can one cope with the reality that Risk A appears to be five times larger than Risk B, but it could well be 214 times larger or B could be 8.6 times larger than A? The path of greatest promise requires three questions to be asked:

- *What is the probability of each error?* If the practical choice involved banning either Substance A or B (or, equivalently, creating a risk of B in the act of eliminating Risk A), then the decision problem above would reduce to this 2 x 2 "consequence table" (note: "A>B" means "A is truly riskier than B"):

Reality	Action	
	Ban A	*Ban B*
A>B	good choice	poor choice
B>A	poor choice	good choice

The first step in enhancing those qualitative descriptions with more tangible consequences is to assign probabilities to each of the two horizontal rows (the axis over which the decision-maker has no control). Here, the probabilities would be about 75% and 25%, respectively (see above).
- *Which of the possible adverse consequences is less tolerable?* In this example, A is more likely to be the larger risk, and it is also more likely that A is much larger than B than vice versa (214 > 8.6). Thus, the "poor choice" in the lower-left corner of the table seems both less likely and less adverse than the "poor choice" in the upper-right corner. In other words, going after A seems sensible because the

consequences of failure (a combination of probability and magnitude) are smaller than they would be if B were addressed instead. However, there are many reasons why a decision-maker might be especially desirous of avoiding certain errors, even if they are less likely and/or less severe (as measured by numbers alone) than other errors. For example, if Substance A was naturally-occurring and B was deliberately added to foods, it might be "rational" to consider precluding a "lower-left" error (allowing the synthetic substance when in fact it is somewhat riskier than the alternative) even though the other mistake (allowing the natural substance when in fact it is much riskier) would have larger "expected consequences." The bottom line of this weighing of likelihoods and consequences in these difficult "Case 3" situations is vexing: *When both a "signal" of relative risk and "noise" are present, one may need to pay less attention to the signal if contained within the noise is a particular unwelcome consequence.*

• *(How) can I improve my prospects?* The final, and perhaps least-appreciated question when faced with choosing among different errors is whether the choice has to be so grim. Sometimes, creative interventions can have a higher expected value and a lower risk of a highly adverse mistake than either (any) of the obvious choices contained in a simplified "consequences table." For example, one might be able to reduce exposures to both Substance A and B without eliminating either one completely. Or, often in the shorter-term, one can reduce the uncertainties without first choosing which risk to reduce. If the consequences of deferring the control choice while the research (uncertainty reduction) proceeds are less onerous than facing an immediate choice would be, the prospects for reducing the "noise" and effectively transforming a Case 3 into a Case 1 may be the most attractive contingency of all. Not all choices can be made less difficult, of course, but (to return one final time to the "lamppost" metaphor) it is almost always worth thinking whether there might be a spare set of keys available to you, rather than continuing to agonize over whether to search either under the lamppost or in the dark.