

8.8

Risk perception and communication

Baruch Fischhoff

Abstract

Public health depends on laypeople's ability to understand the health-related choices that they and their societies face. The study of *risk perception* examines that ability. The study of *risk communication* examines the processes that determine how communication with lay people enhances or degrades their decision-making ability. Although focused on decisions involving risk, that research necessarily considers potential benefits as well, including the benefits of reducing risks (e.g. through medical treatment, lifestyle changes, or improved air quality). Communication is seen as a two-way process. Without listening to people, it is impossible to understand what they know and value, hence impossible to provide them with relevant information in a comprehensible form. The basic science of *behavioural decision research* describes the general processes that find specific expression in risk-related decisions. It provides the conceptual framework, methodology, and theory for this chapter.

Risk perception and communication research is conceptually straightforward. First, characterize the decisions that people face, in sufficiently precise terms to identify the information that is most critical to them. Second, describe people's existing beliefs and values, in sufficiently precise terms to understand their roles in risk-related choices. Third, develop and empirically evaluate communications designed to bridge the critical gaps between what people know and what they need to know, in order to have the best chance of making choices that achieve what they value. These steps are interdependent. For example, descriptive research can reveal unexpected goals, obstacles, and capabilities, forcing revision of the decision analysis; communication failures can force additional research regarding decision-making processes.

Executing a communication research programme requires four kinds of expertise: (a) *subject matter specialists*, (b) *risk and decision analysts*, for characterizing choices and identifying critical information; (c) *behavioural scientists*, for characterizing existing beliefs and values, then designing and empirically evaluating communications; and (d) *communication practitioners*, for executing sustainable programmes. Individuals with each kind of expertise should have final authority in their domain.

Behavioural decision research provides extensive guidance on two topics central to this endeavour. One is identifying potential threats to risk-related decision making, along with an understanding

of the underlying behavioural processes that communications must address. The second is the measurement procedures needed to ensure that people have been properly understood, when creating and evaluating communications. Without proper measurement, it is impossible to assess and address people's information needs. A particular risk is underestimating laypeople's decision-making competence, thereby denying them the opportunity for active participation in health decisions. That risk is aggravated when communications are disseminated without proper evaluation, after which their audience is held responsible for failing to understand content that was neither clear nor relevant. One should no more release untested communications than untested pharmaceuticals. The chapter seeks to reduce those risks, while helping experts to help laypeople to choose wisely.

Introduction

Many health risks arise from deliberate decisions by individuals trying to make choices that balance health and other concerns. Some choices are made as individuals. They include whether to wear bicycle helmets and seatbelts, whether to read and follow safety warnings, whether to buy and use condoms, and how to select and cook food. Other choices are made as citizens. They include whether to protest the siting of hazardous waste incinerators and halfway houses, whether to support fluoridation and "green" candidates, and whether to allow sex education.

Sometimes, single choices have large effects (e.g. buying a safe car, taking a dangerous job, getting pregnant). Sometimes, small effects accumulate over multiple choices (e.g. exercising, avoiding trans fats, wearing seatbelts, using escort services). Sometimes, health-related choices focus on health; sometimes, not (e.g. purchasing homes that require long commutes, choosing friends who exercise regularly, joining religious groups opposed to vaccination).

Making health-related decisions wisely requires understanding the associated risks and benefits. This chapter reviews the research base for characterizing and improving that understanding. Following convention, these are called *risk perception* and *risk communication*, respectively. However, the basic principles also apply to perception and communication regarding the potential benefits of health-related decisions (e.g. lifestyle changes that reduce risks).

Psychologists sometimes reserve ‘perception’ for direct physiological responses to stimuli, using ‘judgement’ for the translation of that response into observable estimates. A currently active research topic is identifying the conditions under which judgement surrenders entirely to perception (Lowenstein *et al.* 2001). Perceptions could prevail either when passions run high or when judgement fails to yield satisfactory choices. This chapter emphasizes judgement, hoping to expand the envelope of deliberative processes in personal and public health decisions.

Inaccurate judgements about risks can hurt people. So can inaccurate beliefs about those judgements. If people’s understanding is overestimated, then they may face impossibly hard choices (e.g. among unfamiliar medical alternatives, without adequate information). If people’s understanding is underestimated, then they may be needlessly denied the right to choose. As a result, the chapter assumes (a) that descriptive statements about people’s beliefs must be disciplined by empirical evidence and (b) that evaluative statements about the adequacy of people’s understanding must be founded on rigorous analysis of what they need to know, in order to make good choices. To these ends, the chapter emphasizes methodological safeguards against misguided assessments.

The next section, ‘Quantitative Assessment’, treats judgements about how big risks are. The following section, ‘Qualitative Assessment’, treats beliefs about the processes that create and control risks, on the basis of which people produce and evaluate quantitative estimates. Both sections address both measurement issues and barriers to understanding. The section on ‘Creating Communications’ provides a structured approach for developing communications about health-related decisions, focused on individuals’ greatest information needs. The ‘Conclusion’ section considers the strategic importance of risk communication in risk management. Access to research on complementary social and emotional processes might begin with Krinsky and Golding (1992), Peters and McCaul (2005), and Slovic (2001).

Quantitative assessment

Estimating risk magnitude

A common complaint among experts is that ‘the public doesn’t realize how small (or large) Risk X is.’ There is empirical evidence demonstrating examples of such biases (Slovic 2001). However, that evidence has typically been collected in settings designed to reveal biases, in order to help researchers study the processes that create them. As a result, the prevalence and magnitude of bias in published studies need not reflect their prevalence and magnitude in life. Generalizing from research decisions to real-world ones requires matching the conditions in each. Looking at one widely cited study in some detail shows how that matching might proceed, while introducing some general principles and results.

Participants

Lichtenstein *et al.* (1978) asked people to estimate the annual number of deaths in the US from 30 causes (e.g. botulism, tornadoes, motor vehicle accidents). These ‘people’ were members of the League of Women Voters and their spouses, making them older than the proverbial college sophomores often studied by psychologists. Age might affect *what* people think, as a result of education and experience. It is less likely to affect *how* they think. Many cognitive processes seem to be widely shared, once people pass

middle adolescence, unless they suffer some impairment (Fischhoff 2008, Reyna & Farley 2006).

One widely shared class of cognitive processes is relying on judgemental *heuristics*, or rules of thumb, when asked to infer unknown quantities (Gilovich *et al.* 2003; Kahneman *et al.* 1982). One well-known heuristic is *availability*, whereby people assess an event’s probability by how easily instances come to mind. Although more available events are often more likely, media coverage (among other things) can make events disproportionately available, inducing biased judgements. How people generate instances, using their memory and imagination, should reflect general cognitive processes. However, the contents of those memories and images should vary with individuals’ experiences. So should their trust in information sources—and attempts to adjust for bias.

Lichtenstein *et al.* (1978) elicited judgements with two *response modes*. One asked people to pick the more frequent of two paired causes of death and, then, to estimate the ratio of their frequencies. The second asked people for the number of deaths. It began by giving the answer for one cause (either electrocution or motor vehicle accidents). That *anchor* was designed to provide a feeling for annual death rates—after pretests found that people often knew little about these statistics. Figure 8.8.1 shows results with the second method.

Results

- (a) Relative risk judgements were consistent, across the two response modes. Risks given higher frequency estimates were typically judged more likely, when paired with risks given lower frequency estimates. The ratios of the direct estimates were similar to the directly estimated ratios. Thus, these people seemed to have an internal ‘scale’ of relative risk, which they expressed consistently even with these unfamiliar tasks.

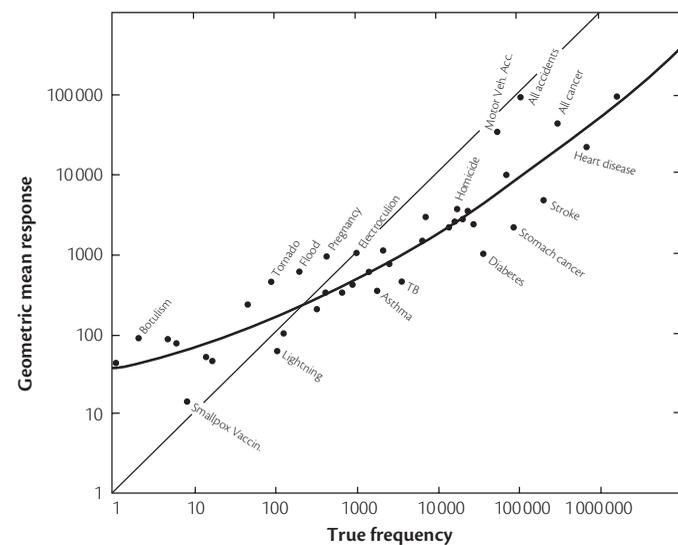


Fig. 8.8.1 Best quadratic fit line to geometric mean judgements of the annual toll from 40 causes of death in the United States, compared to best available statistical estimates.

Source: Lichtenstein *et al.* (1978).

- (b) Absolute risk judgements were affected by the anchor. People told that 50 000 people die annually from auto accidents gave estimates two to five times higher than did people told that 1000 die annually from electrocution. Thus, people seemed to have less feeling for absolute frequency, rendering them sensitive to implicit cues in how questions are posed (Poulton 1989; Schwarz 1999).
- (c) Absolute risk judgements were less dispersed than the corresponding statistical estimates. While the latter varied over six orders of magnitude, individuals' estimates typically ranged over 3–4. That compression could reflect anchoring, if judgements were drawn toward the value that was given for perspective. Overall, people overestimated small frequencies and underestimated large ones. That pattern might change with different anchors. For example, a lower anchor (e.g. botulism deaths) would, likely, reduce (or even eliminate) overestimation of small frequencies, while increasing underestimation of large ones.
- (d) Relative and absolute risk judgements seemed to reflect availability bias. For any statistical frequency, some causes of death consistently received higher estimates (e.g. homicide, tornadoes, flood). These causes were disproportionately reported in the news media and as personal experiences. When told of availability bias, participants could not improve their judgements, consistent with the finding that tracking frequency is an automatic process (e.g. Koriat 1993).

Lichtenstein *et al.* (1978) found some response patterns that were procedure invariant (e.g. relative risk judgements) and some that were not (e.g. absolute estimates). A century of psychophysics research (Poulton 1989) has identified procedural factors that can affect quantitative judgements. Determining their effects in specific settings requires dedicated studies. The practical importance of any bias depends on the decision. Shifting fatality estimates by a factor of 2 might tip some decisions, but not others.

Fischhoff and MacGregor (1983) provide another example of response mode effects. They asked about the chances of dying (in the US), among people afflicted with various maladies (e.g. influenza), in four ways: (a) How many people die out of each 100 000 who get influenza; (b) how many people died out of the 80 million who caught influenza last year; (c) for each person who dies of influenza, how many have it and survive; (d) 800 people died of influenza last year, how many survived? As in Lichtenstein *et al.* (1978), relative risk judgements were consistent across response modes, while absolute estimates varied greatly (over 1–2 orders of magnitude). A second study found that people liked one format (c) much less than the others—and could least well remember statistics reported that way. This format also produced the most discrepant estimates, identifying it as a poor way to elicit or communicate risks.

Evaluative standards

Risk judgements can be evaluated in terms of internal consistency (e.g. whether estimates increase with increasing exposure) or accuracy. Without sound risk estimates (see Chapter 8.7, Toxicology and risk assessment in the analysis and management of environmental risk), one cannot evaluate accuracy. For example, after the 9/11 attacks, some critics claimed that some Americans had

increased their risk level by flying, rather than driving. These claims were based on historical risk statistics. However, at that time, no one knew how safe aviation was (the fleet was grounded), while traffic deaths are disproportionately high among the young, elderly, and drinkers—not the drivers who were shifting transportation modes. Even if the historical statistics were valid, decisions involving risks can reflect other factors as well. Any additional risk of driving might have been justified for someone who was financially strapped by the declining economy or wary of flight delays (with added security hassles). Without understanding all elements of a choice, one cannot judge its reasonableness.

Probability judgements

The sensitivity of quantitative judgements to procedural details might suggest avoiding them, in favour of verbal quantifiers (e.g. likely, rare). Unfortunately, such terms have their own problems, namely, being interpreted differently across people and situations, unless usage norms have evolved (Budescu & Wallsten 1995; Schwarz 1999).

Table 8.8.1 shows verbal and quantitative judgements of seven risks, provided by a fairly homogeneous group (US undergraduates). The quantitative response mode explicitly offered probabilities as low as 0.01 per cent—using a linear scale for 1–100 per cent and expanding 0–1 per cent with log scales from 1:100 to 1:10 000. The qualitative response mode used typical labels (1=very unlikely; 2=unlikely; 3=somewhat unlikely; 4=somewhat likely; 5=likely; 6=very likely). Comparing the two response modes revealed a non-linear relationship: The median probabilities corresponding to the qualitative responses were 0.01 per cent for 'very unlikely,' 0.5 per cent for 'unlikely,' 5 per cent for 'somewhat unlikely,' 25 per cent for 'somewhat likely,' 60 per cent for 'likely,' and 96 per cent for 'very likely'.

Some researchers hesitate to elicit probabilities, lest they exceed laypeople's cognitive capabilities. That hesitation is strengthened by (research and anecdotal) evidence of lay innumeracy. However, even imperfect measures can have value, if their strengths and weaknesses are understood. The research literature on probability

Table 8.8.1 Comparison of numerical verbal and statistical risk estimates ('Please estimate your personal risk to the following events in the next 3 years.')

Risk	Quantitative (Probability) Response (Median %)	Verbal response		Statistical risk Estimate (probability in %)
		Median	Mean	
Electrocution	0.1	1.0	1.67	0.015
Cancer	0.3	2.0	2.09	0.06
Flu	55.0	5.0	4.72	86.2
Car Injury	10.0	3.0	3.38	4.7
Herpes	0.1	1.0	1.73	4.1
AIDS virus/sexual	0.02	1.0	1.41	0.2

Source: Linville *et al.* 1993

elicitation is enormous (O'Hagan *et al.* 2006). Results relevant to public health researchers and practitioners include:

- (a) Numeric probability judgements can be as reliable and acceptable to users as verbal ones. Woloshin *et al.* (1998) found this, comparing linear and log-linear probability scales with verbal ones, for judgements of medical events.
- (b) People often prefer to provide verbal judgements and to receive quantitative ones. Quantitative responses require more effort and entail greater accountability (Erev & Cohen 1990).
- (c) Probability judgements often have good construct validity, correlating sensibly with other variables. For example, Fischhoff *et al.* (2000) found higher probabilities of pregnancy among US teens reporting more sexual activity and high probabilities of being arrested among teens reporting more violent lives.
- (d) Misinformation and mistaken inferences can bias probability judgements. For example, availability can contribute to unwarranted optimism. Our own carefulness (e.g. in avoiding traffic accidents or bad investments) is more 'available' to us than is that of others.
- (e) Probability judgements can be deliberately biased, when people respond strategically. For example, Christensen-Szalanski and Bushyhead (1993) found physicians overestimating the probability of pneumonia, perhaps fearing that the healthcare system would ignore unlikely cases. Probability of precipitation forecasts may show an 'umbrella bias,' overstating chances, to keep people from being caught unprotected (Lichtenstein *et al.* 1982).
- (f) Transient emotions can affect judgements. For example, anger increases optimism, fear the opposite (Lerner & Keltner 2001), with effects large enough to tip close decisions.
- (g) Judgements for the probability of being correct are moderately correlated with actual knowledge. For example, Fischhoff *et al.* (1977) elicited probabilities for successfully choosing the larger of two causes of death (from Lichtenstein *et al.* 1978). Correct choices received higher probabilities. Overall, people were *overconfident* (e.g. with 75 per cent correct choices, when 90 per cent confident). Overconfidence is typical with hard tasks, underconfidence with easy ones.
- (h) Probability judgements can vary by response mode. Differences have been found with odds and probabilities, probabilities and relative frequencies, and judgements of individual or grouped items (Griffin *et al.* 2003).
- (i) Some numerical values are treated specially. For example, people seldom use fractional values (motivating the log-linear scale); when uncertain what to say, people sometimes offer 50, meaning 50–50 and not a numeric probability (Bruine de Bruin *et al.* 2000).
- (j) Probability judgement processes mature by middle adolescence. For example, teens show no greater optimism bias than adults, despite the common belief in adolescent invulnerability (Quadrel *et al.* 1993). Fischhoff *et al.* (2000) found that teens, unlike adults, greatly exaggerate the probability of premature death.
- (k) There are stable individual differences in the ability to use probabilities. They correlate with performance on other tasks,

as well as life outcomes that might reflect decision-making competence (Bruine de Bruin *et al.* 2007).

- (l) The use of probabilities can sometimes be taught. Lichtenstein and Fischhoff (1980) found improvement after a single round of intense feedback.

A test of any measure is its predictive validity. Even though risk decisions often involve choices among options with non-risk consequences, Brewer *et al.* (2007) found that risk judgements alone sometimes have predictive value. In the contexts of smoking (Viscusi 1992) and breast cancer (Black *et al.* 1995), researchers argue that advertising has worked too well, leading to exaggerated fears, a claim supported by unduly high probability judgements (even after deleting apparently non-numeric 50s).

Defining risk

Studies like Lichtenstein *et al.* (1978) measure risk perceptions, if 'risk' means 'chance of death.' However, even among experts, 'risk' has multiple meanings (Fischhoff *et al.* 1984; National Research Council 1996). For those who focus on fatalities, 'risk' might be measured in probability of death, expected life years lost, total deaths, or deaths per person exposed (or per hour of exposure). Each definition entails an ethical position. For example, *life-years lost* places extra weight on deaths of young people, whereas *probability of death* disregards age. Focusing on lost life expectancy increases concern for deaths by injury (e.g. drowning, driving, workplace hazards), relative to deaths from the cumulative effects of chronic illnesses. Adding morbidity and psychological trauma would heighten concern for alcohol and illegal drugs, which can ruin lives without ending them.

Unless such definitional issues are recognized, people can unwittingly speak past one another, when addressing 'risks.' Clarifying definitional issues has been central to risk research. Before reviewing some results, it is worth noting that 'risk' is sometimes used as a discrete variable, treating activities as risky or not. That shorthand says little, without knowing the threshold of concern. Calls for 'safe' products can be unfairly ridiculed, by treating them as demanding zero-risk. Critics of cost-benefit analysis have offered various *precautionary principles*, for avoiding risks too uncertain to countenance. However, they have limited use until one specifies the threshold of concern and procedures for assessing compliance (DeKay *et al.* 2002).

Catastrophic potential

One early risk perception study asked experts and laypeople to estimate the US 'risk of death' from 30 activities and technologies (Slovic *et al.* 1979). These judgements correlated more strongly with statistical estimates of average-year fatalities for experts than for laypeople. However, when asked to estimate average-year fatalities, laypeople responded like experts. Inspection suggested that laypeople interpreted 'risk of death' to include catastrophic potential, reflecting the expected deaths in non-average years. If so, then experts and laypeople agreed about routine deaths (which have relatively good scientific estimates) and disagreed about possible anomalies (for which the science is much weaker). Such potentially reasonable disagreement would be obscured by the casual assumption that any disagreement between experts and laypeople reflects lay ignorance (National Research Council 1989).

The moral principle underlying this definitional disagreement could mean valuing deaths more when lost at once than when

lost individually. Slovic *et al.* (1984) found, however, that catastrophic potential worries people because it suggests technologies that might spin out of control. An aversion to deep uncertainty should be less controversial.

Dimensions of risk

Beginning with Starr (1969), many features, like uncertainty and catastrophic potential, have been suggested as affecting definitions of risk. In order to reduce the set to manageable size, Fischhoff *et al.* (1978) had members of a liberal civic organization rate 30 hazards on nine such features. Factor analysis on mean ratings identified two *dimensions*, accounting for 78 per cent of the variance. Similar patterns emerged with students, members of a conservative civic organization, members of a liberal women's organization, and risk experts. Figure 8.8.2 plots factor scores within the common factor space for these four groups.

Hazards high on the vertical factor (e.g. food colouring, pesticides) were rated as new, unknown, and involuntary, with delayed effects. Hazards high on the horizontal factor (e.g. nuclear power, commercial aviation) were rated as fatal to many people, if things go wrong. The factors were labelled *unknown* and *dread*, respectively. They might be seen as capturing the cognitive and emotional bases of people's concern, respectively.

Many studies, using this 'psychometric paradigm' have found roughly similar dimensions, despite differing elicitation mode, scaling techniques, items, and participants (Slovic 2001). When a third dimension emerges, it appears to reflect the scope of the

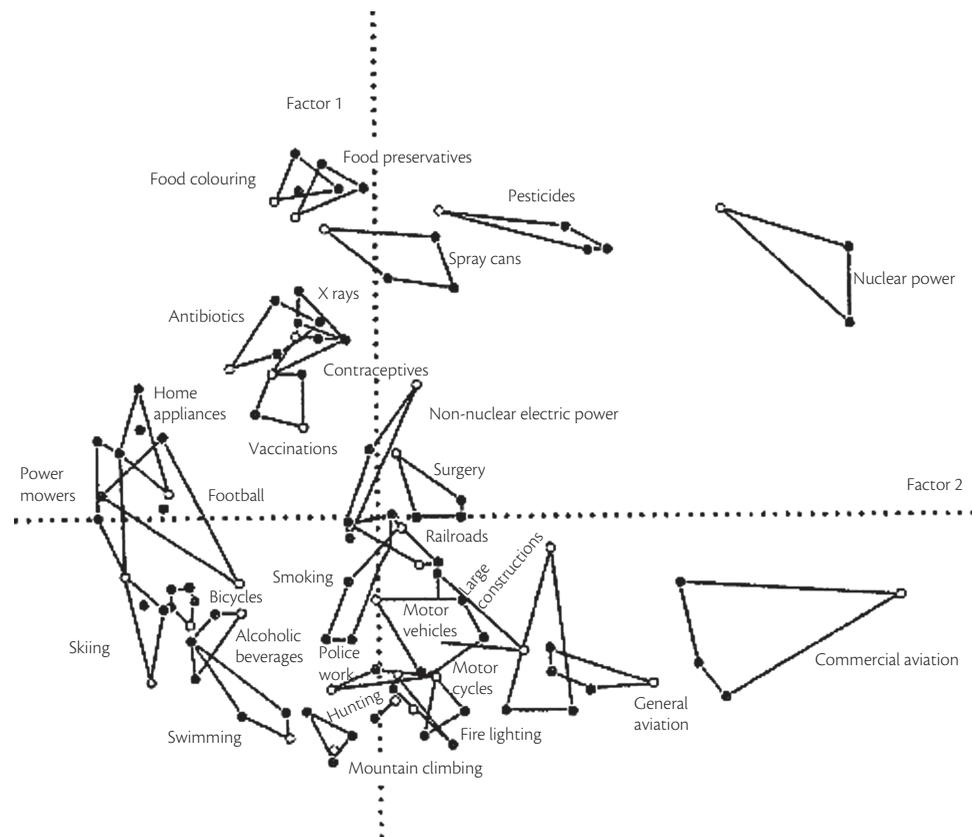
threat, labelled *catastrophic potential*. Hazards' position in the space correlates with attitudes toward them, such as the desired stringency of regulation. Analyses of mean responses are best suited to predicting aggregate (societal) responses. Individual differences have also been studied (e.g. Vlek & Stallen 1981).

Risk comparisons

The multidimensionality of risk means that hazards similar on some dimensions may still evoke (and deserve) quite different responses. This fact is neglected in appeals to accept a risk because one has accepted another risk with some similarities (Fischhoff *et al.* 1984). Such risk comparisons sometimes present many hazards, in quantities posing equal statistical risks (e.g. both a tablespoonful of peanut butter and 50 years living by a nuclear power plant create a one-in-a-million risk of premature death). Box 8.8.1 shows such potential flaws in risk comparisons.

One way to improve the legitimacy of risk comparisons is to involve users in setting them. Following this strategy, the US Environmental Protection Agency (1993) promoted some 50 regional, state, and national risk-ranking exercises, in which citizens deliberated priorities on dimensions of their choosing, supported by technical staff providing relevant analyses. Participants' freedom to choose dimensions made individual exercises more relevant, while reducing comparability across exercises. Florig *et al.* (2001) developed a method for standardizing such comparisons, based on the risk dimensions research (Table 8.8.2). The UK government has endorsed a variant (HM Treasury 2005).

Fig. 8.8.2 Location of 30 hazards within the two-factor space obtained from League of Women Voters, student, active Club, and expert groups. Respondents evaluated each activity or technology on each of nine features. Ratings were subjected to principal components factor analysis, with a varimax rotation. Connected lines join or enclose the loci of four group points for each hazard. Open circles represent data from the expert group. Unattached points represent groups that fall within the triangle created by the other three groups. Source: Slovic *et al.* 1985.



Box 8.8.1 Risk comparisons

One ... legitimate purpose [for risk comparisons] is giving recipients an intuitive feeling for just how large a risk is by comparing it with another, otherwise similar, risk that recipients understand. For example, roughly one American in a million dies from lightning in an average year (NOAA 1995). 'As likely as being hit by lightning' would be a relevant and useful comparison for someone who has an accurate intuitive feeling for the probability of being hit by lightning, faces roughly that 'average' risk, and considers the comparison risk to be like death by lightning in all important respects. It is not hard to imagine each of these conditions failing, rendering the comparisons irrelevant or harmful:

- (a) Lightning deaths are so vivid and newsworthy that they might be overestimated relative to other, equally probable events. But 'being struck by lightning' is an iconic very-low-probability risk, meaning that it might be underestimated. Where either occurs, the comparison will mislead.
- (b) Individual Americans face different risks from lightning. For example, they are, on the average, much higher for golfers than for nursing-home residents. A blanket statement would mislead readers who did not think about this variability and what their risk is relative to that of the average American.
- (c) Death by lightning has distinctive properties. It is sometimes immediate, sometimes preceded by painful suffering. It can leave victims and their survivors unprepared. It offers some possibility of risk reduction, which people may understand to some degree. It poses an acute threat at some very limited times but typically no threat at all. Each of those properties may lead people to judge them differently—and undermine the relevance of comparisons with risks having different properties.
- (d) It is often assumed that the risks being used for comparison are widely considered acceptable at their present levels. The risks may be accepted in the trivial sense that people are, in fact, living with them. But that does not make them acceptable in the sense that people believe that they are as low as they should or could be ...

The second conceivable use of risk comparisons is to facilitate making consistent decisions regarding different risks. Other things being equal, one would want similar risks from different sources to be treated the same. However, many things might need to be held equal, including the various properties of risks ... that might make people want to treat them differently despite similarity in one dimension ...

The same risk may be acceptable in one setting but not another if the associated benefits are different (for example, being struck by lightning while golfing or working on a road crew). Even when making voluntary decisions, people do not accept risks in isolation but in the context of the associated benefits. As a result, acceptable risk is a misnomer except as shorthand for a voluntarily assumed risk accompanied by acceptable benefits.

Source: National Research Council 2006.

Table 8.8.2 A standard multi-dimensional representation of risks

Number of people affected	Degree of environmental impact	Knowledge	Dread
Annual expected number of fatalities: 0– 450 –600 (10% chance of zero)	Area affected by ecosystem stress or change 50 km ²	Degree to which impacts are delayed 1–10 years	Catastrophic potential 1000 times expected annual fatalities
Annual expected number of person-years lost 0– 9000 –18 000 (10% chance of zero)	Magnitude of environmental impact modest (15% chance of large)	Quality of scientific understanding medium	Outcome equity medium (ratio = 6)

Source: Adapted from stimuli used in research reported by Willis *et al.* 2005

Qualitative assessment

Event definitions

Once defined, 'risk' can be estimated. That requires specifying the conditions for its observation. For example, when estimating the 'risk' of pregnancy, conditions include the frequency and timing of intercourse, contraceptives used, and partners' physical state. Unless laypeople are given similar detail, when asked to judge risks, they cannot convey their beliefs. Unfortunately, many survey questions leave respondents guessing at their meaning. Consider, for example, 'How likely do you think it is that a person will get the AIDS virus from sharing plates, forks, or glasses with someone who had AIDS?' US college students answered this question (taken from a prominent national survey), then were asked what they had inferred about the kind and amount of sharing. They generally agreed about the kind, with 82 per cent choosing 'sharing during a meal' from a set of options. However, they disagreed about the frequency (a single occasion, 39 per cent; several occasions, 20 per cent; routinely, 28 per cent; uncertain, 12 per cent) (Fischhoff 1996). Thus, they were, effectively, answering different questions, whose meaning could only be guessed by those hearing their responses.

Laypeople are, similarly, left guessing when experts communicate risks ambiguously (Fischhoff 1994). For example, McIntyre and West (1992) found that teens knew that 'safe sex' was important, but disagreed about what it entailed. Downs *et al.* (2004b) found that teens interpret 'it can only take once' as meaning that they will get pregnant after having sex once. If they do not, some infer that they are infertile, encouraging unsafe sex. Murphy *et al.* (1980) found people divided over whether '70 per cent chance of rain' referred to (a) the area receiving rain, (b) the time it would rain, (c) the chance of some rain anywhere, or (d) the chance of some rain at the weather station (the correct answer). Fischhoff (2005a) describes procedures for improving and evaluating event definitions, so that experts and laypeople understand one another well enough to be talking about the same thing.

Supplying details

The details that people infer, when given ambiguous risk questions or messages, reveal their intuitive theories. For example, when teens thought aloud while judging the probabilities of ambiguous events based on survey questions (e.g. having an accident after drinking and driving, getting AIDS through sex), they typically noticed many unstated details, usually focusing on ones that would affect scientific risk estimates (Fischhoff 1994). For example, they wondered about the 'dose' of most risks (e.g. the amount of drinking and driving), which was not stated in any of the questions. An exception was not thinking about the amount of sex when judging the risks of pregnancy and HIV transmission. Teens seemed to believe that an individual is either vulnerable or not (consistent with Downs *et al.* 2004b and other results reported there). Sometimes they considered variables that were not clearly related to risk, such as how well partners know one another. In an interactive DVD that reduced adolescent sexual risks, Downs *et al.* (2004a) addressed how partners could fail to self-diagnose STIs.

Cumulative risk— A case in point

There is no full substitute for directly studying the beliefs that people bring to and take away from risk messages. However, the research literature provides a basis for anticipating those beliefs.

For example, the optimism bias is so widespread with events where some personal control seems feasible that one can assume that people see themselves as facing less risk when told (or asked) about others' risk. Similarly, teens' insensitivity to the amount of sex, when judging STI risks, reflects a well-known insensitivity to how risks accumulate over repeated exposure. Thus, people cannot be expected to infer the accident risk from repeatedly driving without a seatbelt (Slovic *et al.* 1978) or the pregnancy risk from having sex with generally effective contraceptives (Shaklee & Fischhoff 1990). One corollary of this insensitivity is not realizing the cumulative impact of small differences in single-exposure risks (e.g. slightly better contraceptives, wearing a seatbelt). People similarly underestimate exponential growth (e.g. Frederick 2005). Some people have difficulty with the mental arithmetic; others see no risk-exposure relationship.

For example, Linville *et al.* (1993) had college students judge the probability of transmission from an HIV-positive man to a woman from 1, 10, or 100 cases of protected sex. For one case, the median estimate was 0.10, much higher than then-current public health estimates—despite using a log-linear response mode that facilitated making very low probability judgements. The median estimate for 100 contacts was 0.25, a more accurate estimate, but one that reveals typical under-accumulation. Studies that asked about just one or just 100 exposures would reveal very different pictures of lay beliefs. Conversely, risk messages could convey very different pictures if they reported risks of just one or just 100 exposures. A complete picture requires providing and asking about both exposures.

Mental models of risk processes

The role of mental models

As mentioned, when people lack explicit information about the magnitude of risks (and benefits), they must infer it. Judgemental heuristics, like availability, provide one class of inferential rules, deriving quantitative estimates from experience. Other inferences are derived from people's *mental models* of the processes creating and controlling risks. Those intuitive theories serve other functions, beyond providing estimates useful for (relatively) well-formulated decisions. They allow people to follow issues in the news media, participate in discussions, feel competent to make decisions, and generate choice options.

The term 'mental model' is often applied to intuitive theories that are well enough elaborated to generate predictions or explanations in diverse circumstances. Mental models have a long history in psychology, studied for topics as diverse as how people understand physical processes, international tensions, complex equipment, energy conservation, interpersonal relations, and drug effects (Ericsson & Simon 1993).

If mental models contain 'bugs,' they can produce erroneous conclusions, even for otherwise well-informed individuals. For example, not realizing how quickly the risks of pregnancy and STIs accumulate with additional sex acts could undermine other knowledge. Bostrom *et al.* (1992) found that many people know that radon is a colourless, odourless, radioactive gas. Unfortunately, people also associate radioactivity with permanent contamination. However, this (widely publicized) property of high-level waste is not shared by radon, whose relevant byproducts have short half-lives. Not realizing this, homeowners might not bother to test, believing that they could do nothing if they found a problem and not knowing that the rapid decay means rapid energy release.

Different methods have evolved for eliciting mental models of different processes. With health risks, the initial measurement challenge is determining the factors that people consider relevant. Morgan *et al.* (2001) offer a strategy that has been used for varied risks. It begins by creating a formal model, summarizing scientific knowledge of the processes affecting risk levels. It should be sufficiently precise in specifying its variables and relationships so that quantitative predictions could be computed, were its data needs met (Fischhoff *et al.* 2006). A common formalism is the influence diagram (Howard 1989). Figure 8.8.3 shows part of an influence diagram for radon. An arrow means that the value of the variable at its head depends on the value of the variable at its tail. Thus, the lungs' particle clearance rate depends on individuals' smoking history. Other examples include STIs (Fischhoff *et al.* 1998), breast implants (Byram *et al.* 2001), sexual assault (Fischhoff 1992), Lyme disease, falls, sexual assault, breast cancer, vaccination, infectious disease, and nuclear energy sources in space (Downs *et al.* 2008; Fischhoff 2005b; Morgan *et al.* 2001).

The research continues with open-ended individual interviews, structured around the model. Interviews begin very generally, asking what respondents know about the topic, then requesting elaboration on each issue they raise. The tone is non-judgemental, seeking to understand respondents' perspectives, not evaluate them. Interviews proceed to ask about exposure, effect, and mitigation issues—topics so basic that mentioning them would correct

an oversight, rather than introduce foreign concepts. Once these general topics have been exhausted, more specific issues are raised (e.g. 'How does the amount of sex [or number of partners] affect HIV risk?'; 'What does 'safe sex' mean?'). Another technique is having people think aloud while sorting diverse photographs by their relevance, hoping to evoke neglected topics, without inducing improvised ones. For example, seeing a supermarket produce counter led some respondents to say that radon in the air or soil might contaminate plants (Bostrom *et al.* 1992).

Once transcribed, interviews are coded into the expert model of the risk, adding elements raised by respondents. Those additions might be errors or reflections of lay expertise (e.g. about their own behaviour, unstudied side effects, or how equipment really works). Once mapped, lay beliefs can be analysed in terms of their accuracy, relevance, specificity, and focus. Coding for accuracy can reveal beliefs that are correct and relevant, clearly wrong, too vague to evaluate, correct but peripheral (suggesting misplaced attention), and broadly relevant (e.g. radon is a gas). Bostrom *et al.* (1992) interviewed individuals drawn from civic groups. Most knew that radon is a gas (88 per cent), which concentrates indoors (92 per cent), is detectable with a test kit (96 per cent), comes from underground (83 per cent), and can cause cancer (63 per cent). However, many also believed erroneously that radon affects plants (58 per cent), contaminates blood (38 per cent), and causes breast cancer (29 per cent). Few (8 per cent) mentioned that radon decays.

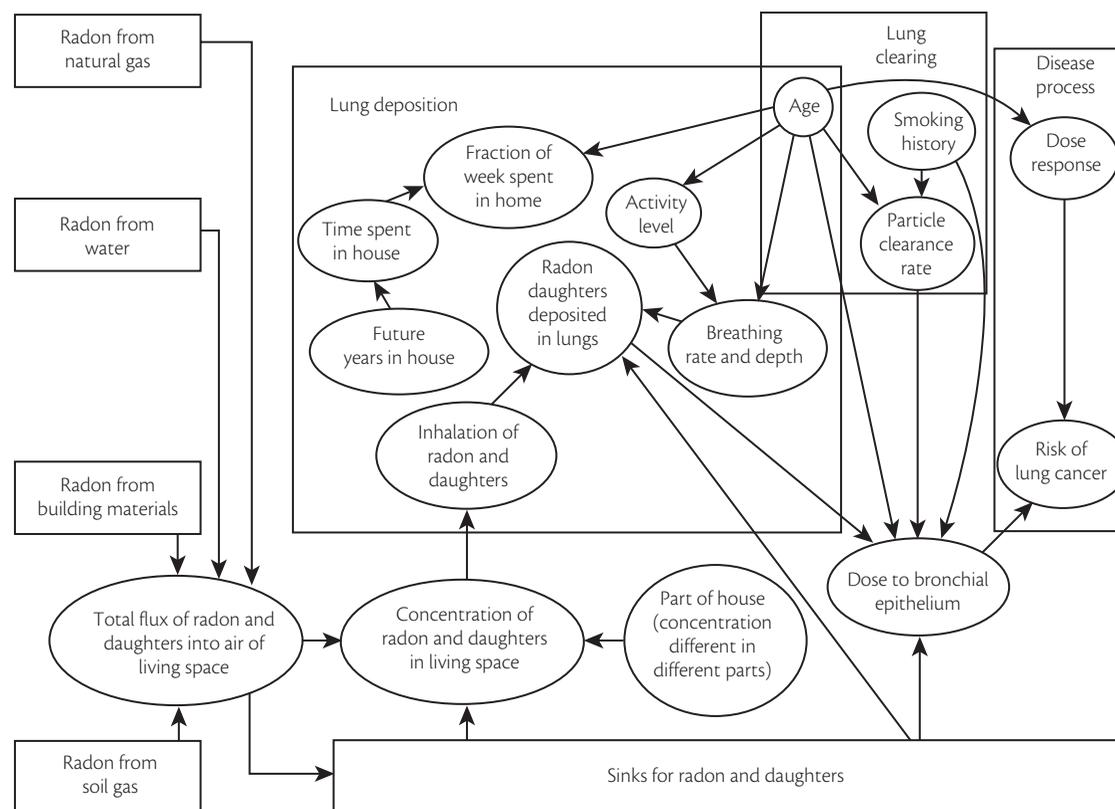


Fig. 8.8.3 Expert influence diagram for health effects of radon (in a home with a crawl space). This diagram was used as a standard and as an organizing device to characterize the content of lay mental models. Source: Morgan *et al.* 1992.

The robustness of these interview results was examined (and generally confirmed) in larger samples, using structured questionnaires that reflected the content and wording of the interview.

From risk beliefs to risk decisions

As mentioned, reasonable decisions should reflect all the outcomes possibly arising from the possible choices. That context is also needed to assess the adequacy of risk perceptions. Some decisions require precision, others just a rough idea. For example, von Winterfeldt and Edwards (1986) showed that decisions with continuous options (e.g. invest \$X) are often insensitive to imprecision in individual input variables (i.e. probabilities, values)—although multiple, correlated errors have cumulative effects. Dawes *et al.* (1989) showed that choices with discrete outcomes (e.g. graduate candidates) are often insensitive to how predictors are weighted, when using simple linear (weighted sum) models. As a result, any model that considers the probability and magnitude of consequences should have some success in predicting behaviour, if applied by researchers familiar with the topics on people's minds. On the other hand, because many such models will do reasonably well, it is difficult to distinguish among them or to gain insight into underlying processes.

Feather (1992) provides a general account of such *expectancy-value* (probability-consequence) models, in which decisions are predicted by multiplying ratings of the likelihood and of the (un)desirability of seemingly relevant consequences. The health-belief model and theory of reasoned action fall into this general category. For example, Bauman (1980) had seventh graders rate 54 possible consequences of using marijuana, in terms of their importance, likelihood, and valence (positive or negative). A 'utility structure index,' computed from these three judgements, predicted about 20 per cent of the variance in subjects' reported marijuana usage.

Just as semi-structured, open-ended interviews can elicit mental models of risk processes, they can also elicit mental models of risk decisions. The template for studying these perceptions is a *decision tree*, that includes the options, relevant outcomes, and uncertain events linking the two. Figures 8.8.4 and 8.8.5 shows two simple decision trees, one for drinking and driving and one for taking the dietary supplement, saw palmetto.

Beyth-Marom *et al.* (1993) had teens and parents from low-risk settings (e.g. sports teams, service clubs) produce possible consequences

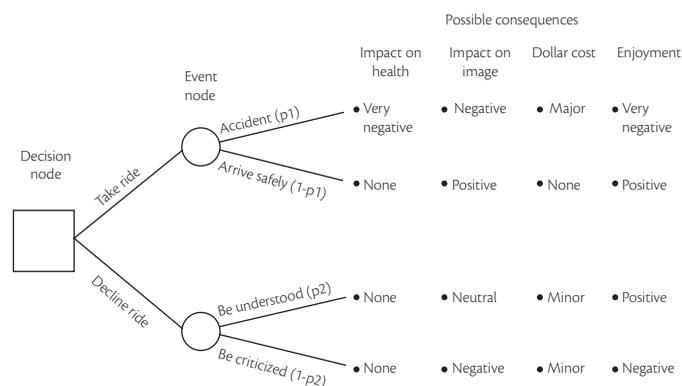


Fig. 8.8.4 A simple decision tree for whether to ride with friends who have been drinking
Source: Fischhoff & Quadrel 1991.

of accepting or rejecting a risky option (e.g. drinking and driving, smoking marijuana). Although accepting and rejecting are formally complementary, they can stimulate different thoughts (Schwarz 1999). Here, accepting the risky options evoked more consequences (suggesting that action was more evocative), a higher ratio of bad to good consequences (suggesting that its risks were more available), and fewer references to social consequences. Making some choices repeatedly evoked different consequences than did doing them once (e.g. more social reactions for repeatedly 'accepting an offer to smoke marijuana at a party'). Teens and parents responded similarly, except that parents mentioned more long-term consequences (e.g. ruining career prospects). These different conceptualizations would be hidden with structured surveys, eliciting ratings of fixed, predetermined consequences.

Fischhoff (1996) reports a study imposing even less structure, with teens describing three difficult personal decisions, in their own terms. These descriptions were coded in terms of their content (what choices trouble teens) and structure (how they were formulated). None of the 105 teens mentioned a choice about drinking-and-driving, while many described drinking decisions. Few decisions had option structures as complicated as Fig.8.8.4. Rather, most had but one option (e.g. whether to attend a party with drinking). Judging by Beyth-Marom *et al.*'s (1993) results, teens looking at that one option saw a different decision than teens focusing on another possible option (e.g. not attending the party, going somewhere else) or multiple ones. Experimental research has found that the opportunity costs (foregone benefits) of neglected options are less visible than their direct consequences (Thaler 1991). For example, the direct risks of vaccinating children can loom disproportionately larger than the indirect risks of not vaccinating them (Ritov & Baron 1990).

Different methods have different strengths and weaknesses (Ericsson & Simon 1993). Structured ones risk suppressing important behaviours, by affording them no expression, while unwittingly misleading respondents, by not allowing confusion to emerge. Unstructured ones risk manipulating behaviour, through the interaction between researcher and respondent, while lacking standardization. Together, these methods can provide a rounded picture, especially when coordinated by normative analysis (which also allows reliable coding of responses to unstructured methods).

With complex, unfamiliar risks, there may be no substitute for interactive procedures, helping people to engage the topic (Fischhoff 2005a). The goal is simulating real-life learning in a beneficent world, trying to deepen their perceptions, without biasing them. That would entail one-on-one interaction, unless group experience were natural. It would rarely entail focus groups, except for generating ideas. The inventor of focus groups, Robert Merton (1987) rejected them as sources of evidence, given the unnatural discourse of even the best-moderated group, the difficulty of hearing individuals out, and the impressionistic coding of contributions. He preferred focused interviews, akin to mental models interviews without the normative analysis.

Creating communications

Selecting information

Communication design should begin by selecting its content. The gold standard is a normative analysis, identifying the information most relevant to the specific choices facing recipients. In practice,

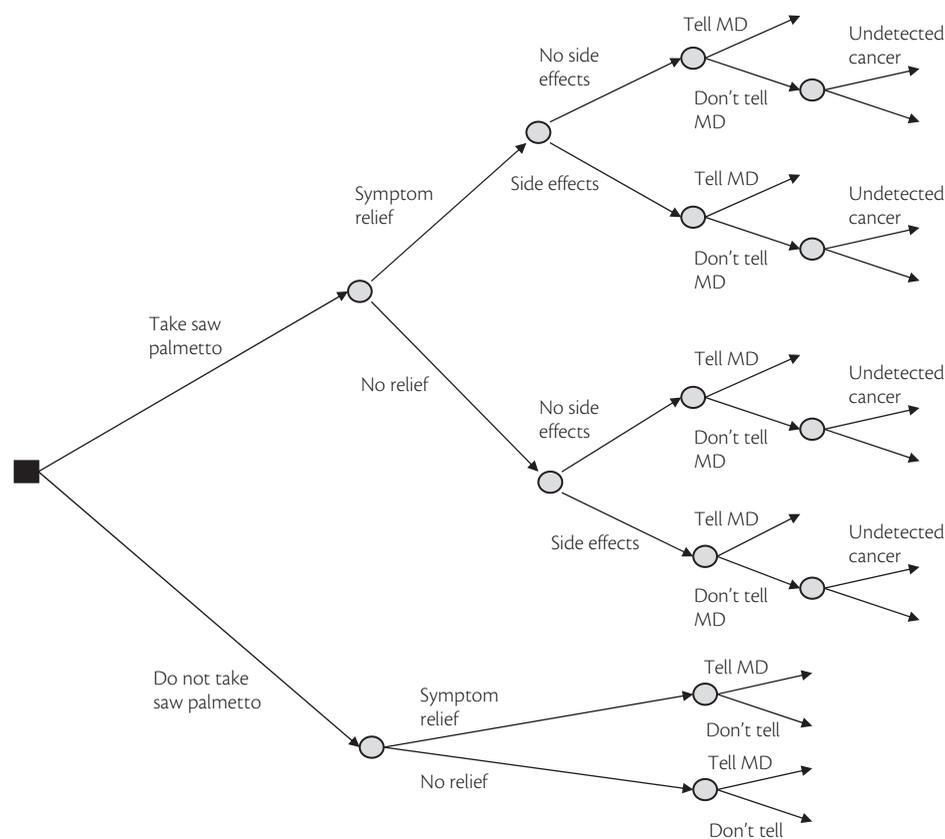


Fig. 8.8.5 A simple decision tree for whether to take saw palmetto for benign prostatic hyperplasia. Source: Eggers & Fischhoff 2004.

the process is disturbingly ad hoc, with self-appointed experts intuiting ‘what people ought to know.’ Not only can poorly chosen information waste recipients’ time; it can also erode their faith in experts (and the institutions employing them). It has opportunity costs, taking the place of needed content. It allows recipients to be judged unfairly, if they seem unable to learn, when they are actually denied a meaningful chance or are uninterested in information that seems irrelevant to them, even if experts deem it important. The Institute of Medicine’s landmark report, *Confronting AIDS* (1986), despaired over a survey finding that only 41 per cent of the public knew that a virus caused AIDS. Yet, one might ask what practical value that information has (and what ‘a virus’ meant for those who answered correctly). Florig and Fischhoff (2007) find that an official list of emergency provisions is outside many individuals’ budget. Their analysis also shows that even those who can afford the stockpile might see it as not worth the cost, given the minuscule probability of proving effective.

Here are three approaches to determining what to say:

- ◆ *Complete mental models, by bridging the gaps between expert and lay mental models.* That could mean adding missing concepts, correcting mistakes, strengthening correct beliefs, and de-emphasizing peripheral ones. Following the method given above: (a) Define the universe of relevant expert knowledge; (b) elicit current beliefs; and (c) assess the centrality of imperfect beliefs.
- ◆ *Ensure appropriate confidence in beliefs.* The most dangerous beliefs are those held with too great or too little confidence.

The appropriateness of confidence can be assessed by comparing judged probabilities of being correct with actual ones. Then focus communication on cases where overconfidence could cause poor choices or under-confidence could prevent sound ones. Routinely communicating how well facts are known might improve the appropriateness of recipients’ confidence. For example, a meta-analysis (Fortney 1988) concluded, with great confidence, that oral contraceptives may increase a non-smoking woman’s life expectancy by up to 4 days and decrease it by up to 80 days. Moreover, the existing research base was so large that no conceivable study could materially change those bounds. That might be enough for some choices. Probability of precipitation forecasts (Murphy *et al.* 1980) show the value of providing information about definitiveness.

- ◆ *Provide information in the order of its expected impact on decisions.* Value-of-information analysis determines a fact’s expected contribution to decision outcomes. It can create a ‘supply curve,’ prioritizing facts by their value. For example, Merz *et al.* (1993) examined the potential risks of carotid endarterectomy. Scraping out the artery to the head can reduce stroke risk, but also cause many problems. The research created a population of hypothetical patients, varying in their physical condition and health preferences, all of whom would want the procedure, were there no side effects. The analysis found that only three possible side effects (death, stroke, facial paralysis) were likely and severe enough that considering them should change many decisions. Although nothing should be hidden, communications should get these few facts across. Arguably, the materiality standard in

medical informed consent could be operationalized this way. Value-of-information analysis can also set priorities for applied research, by clarifying its contribution to decision making.

Value-of-information analysis might be used to identify focal facts. Calibration analysis might be used to identify surprising facts, capable of grabbing recipients' attention and changing their behaviour. A mental model analysis might be used to structure explanatory materials.

Formatting information

Once selected, information must be presented. Reimer and Van Nevel (1999) and Wogalter (2006) provide points of access to research regarding alternative displays. For example, text comprehension research finds that (a) comprehension improves when text has a clear structure, corresponding to recipients' intuitive representation; (b) information leading a clear hierarchy is remembered best; and (c) readers benefit from *adjunct aids*, such as highlighting, advanced organizers (showing what to expect), and summaries.

As elsewhere, the magnitude of these effects in specific settings must be studied empirically. Riley *et al.* (2001) developed a general method for evaluating the adequacy of communications, demonstrated with methylene chloride-based paint stripper. It begins by estimating the risks to users taking different precautionary measures (using an inhalation-uptake model for peak and total chemical exposure). It then evaluates product labels making different assumptions about how users read and follow their content. Possible reading patterns include reading the first five items, reading instructions only, reading highlighted material only, and reading every word. Their prevalence can be assumed or observed in actual use. How well people follow what they read can be estimated from mental models interviews or observation studies. The study found widely varying risk for the same product, depending on their labels. Some provided critical, useful precautionary information for all readers, while some provided it for just some readers (e.g. those looking for warnings), and others hardly provided it at all.

Evaluating communications

However sound their theoretical foundations, communications must be empirically evaluated (National Research Council 1989; Slovic 2001). One should no more release an untested health communication than an untested drug. Indeed, communications could be seen as part of drugs, shaping how they are chosen, used, and monitored. Arguably, that evidence should be part of regulatory filings for approval and part of post-licensing surveillance, especially for drugs available over-the-counter or used off-label. A minimum standard is that recipients understand the content when initially read. More ambitious tests include remembering it later, demonstrating active mastery by making inferences in novel situations, and reaching personally optimal choices.

Evaluating what people learn from communications faces the same challenges as measuring their current risk perceptions. One wants to avoid restricting the expression of non-expert beliefs, suppressing inconsistent beliefs, and changing beliefs through cues embedded in how questions and answers are phrased. Table 8.8.3 summarizes approaches to reader-based evaluation. Open-ended interviews are the best way to reduce these threats. However, performing them to scientific standards is labor intensive. It entails

conducting, transcribing, and coding interviews, with suitable reliability checks, in addition to producing the expert model needed for their evaluation.

The stakes riding on many risk communications (for both the communicator and the recipient) should justify that investment. When time and financial resources are unavailable, almost any data collection is better than none. Even a few open-ended, one-on-one interviews might catch incomprehensible or offensive material, and produce results that might motivate more intense data collection. It is depressing how often even rudimentary evaluation is missing. Amateurish, unscientific communications can be worse than nothing, by creating the misleading impression that the problem has been addressed. Scientific evaluation requires methods taken from the research literature, which has established their strengths and weaknesses, then applied to a standard that could withstand scientific peer review (even if specific applications lack the general interest needed to merit publication).

The methods described here are suited for both *persuasive communication*, designed to manipulate individuals to act in ways determined by the communicator, and *non-persuasive communication*, designed to help individuals identify actions in their own best interest. The two approaches could converge, if persuasive communicators establish that their goal is the action that well-informed individuals should pursue. That action should, of course, reflect all decision outcomes and not just the ones that interest the communicator. For example, Bostrom *et al.* (1992) found people who did not test for radon in order to avoid creating evidence that could complicate selling their home. Fischhoff (1992) reports on the conflicting advice regarding how to reduce the risk of sexual assault, apparently reflecting advisors' different views on women's goals and the division of responsibility among women and society, as well as on the effectiveness of possible actions. Slovic and Fischhoff (1983) describe how reasonable individuals may 'defeat' safety measures by gaining more benefit from a product (e.g. driving faster with a car that handles better)—even if that does not satisfy the safety engineers.

Managing communication processes

In order to communicate effectively organizations require four kinds of expertise:

- a. *Subject matter specialists*, who can identify the processes creating and controlling risks (and benefits).
- b. *Risk and decision analysts*, who can estimate the risks (and benefits) most pertinent to decision makers (based on subject matter specialists' knowledge),
- c. *Behavioural scientists*, who can assess decision makers' beliefs and goals, guide the formulation of communications, and evaluate their success.
- d. *Communication practitioners*, who can manage communication products and channels, getting messages to audiences and feedback from them.

These experts' work must be coordinated, so that they play appropriate roles. For example, behavioural scientists should not revise text (for improved comprehensibility), without having subject matter specialists check that the content has not been changed; subject matter specialists should not slant the facts according to

Table 8.8.3 Data collection options for reader-based evaluations of risk communications

	Strengths	Weaknesses
Concurrent Think-aloud protocol	Protocols identify specific problems with text content and organization; can produce surprises	Costly, time-consuming; difficult to analyse; samples usually small
Retrospective Open-ended	Least reactive—avoids structuring answers for respondents	Coding scheme necessary—data potentially difficult to analyse
Interview	Identifies how reader structures knowledge, is less reactive than most methods	Costly, time-consuming; samples usually small
Short questions, recall	Measures what 'sticks' in readers' minds; can measure how readers assign importance	May not elicit information used in actual decision-making; responses driven by context; difficult to analyse
Problem solving (scenarios)	Elicits decision-making information and strategies	Frames problems for respondents—may be reactive
Closed-ended	Data structured, hence easier and cheaper to collect and analyse; large samples more feasible	Potentially reactive—may misrepresent respondents' knowledge and attitudes
Knowledge tests (true-false, multiple-choice)	Can verify specific misconceptions and beliefs; data readily comparable	Costly; difficult to design valid questions and response scales

Source: Bostrom *et al.* 1994

their pet theories of how the public needs to be alarmed or calmed. Without qualified experts, these roles will be filled by amateurs, imperiling the communicating organization and its public.

Conclusion

Effective risk communication is essential to managing risks in socially acceptable ways. Without it, individuals are denied the best chances of making sound choices, before, during, and after problems arise. As a result, they may suffer avoidable injuries, along with the insult of feeling that the authorities have let them down, by failing to create and disseminate the information that they need, in a timely, comprehensible way. One should no more expose individuals to an untested risk communication than to an untested drug.

Effective risk communication focuses on the decisions that people face. Without that focus, one cannot know what information they need. Sound risk management requires not only communicating that information, but also creating it, both through risk analyses, summarizing existing research (Chapter 8.7, Toxicology and risk assessment in the analysis and management of environmental risk), and new research creating the basis for risk analyses (most other chapters in the textbook).

As a result, risk communication is not just an afterthought, letting the public know what the experts and authorities have decided. Rather, it is central to risk management, providing a disciplined way of communicating decision makers' needs to policy makers. It begins the process by analysing risks from decision makers' perspectives, giving formal representation to the situations they face and the help they need. That analysis also helps to ensure that individuals are judged fairly, when evaluating their risk perceptions and decisions.

This chapter has focused on measurement. In part, that is because the research relevant to each risk domain entails details specific to those decisions. Separate chapters could cover the details for diabetes, drugs, driving, etc. In part, that is because good measurement is essential to good science. Moreover, the methods are sufficiently

general and well understood that they could be applied in any domain. Given a well-characterized decision or risk, it is relatively straightforward, if technically demanding, to assess lay (or expert) perceptions.

Given good measurement of risk and benefit perceptions, risk-related choices can often be roughly predicted with a simple linear model (Dawes *et al.* 1989). More precise prediction requires more detailed understanding of the processes shaping these beliefs, as well as an understanding of the emotional, social, economic, and other processes impinging on specific decisions. Prediction may not be that important, if the public health goal is helping people to make the best choices, given their circumstances—or to empower them to change those circumstances.

Meeting the risk perception and communication challenge requires coordinating the activities of four kinds of experts: Subject matter specialists, risk and decision analysts, behavioural scientists, and communication practitioners. Without them, risk communication will be mismanaged. With them, organizations will have access to the reservoirs of knowledge in their respective disciplines. For example, behavioural scientists trained in decision making will also know something about cognitive, health, and social psychology, as well as whom to call for more.

Although innovative research continues in these constituent fields, the initial challenge for risk communication is taking advantage of what is known already. There is no good reason for the measurement of risk perceptions and the evaluation of risk communications to use less than the readily available methods described here. There is no good reason to ignore well-established results, such as the multidimensional character of 'risk,' the problems with verbal quantifiers, and the need to help people to understand how risks mount up through repeated exposure. Ad hoc communications might reflect sound intuitions, but they deserve less trust than scientifically developed ones.

By definition, better risk communication should help its recipients to make better choices. It need not make the communicators'

lives easier—recipients may discover bonafide disagreements with the communicators and their institutions. What it should do is avoid conflicts due to misunderstanding, increasing the light-to-heat ratio in risk management, leading to fewer but better conflicts (Fischhoff 1995).

Summary

- ◆ Risk communication is central to public health. Without it, individuals are denied the opportunity to make the best possible choices for themselves, their families, and their society.
- ◆ Scientifically sound risk communication requires (a) explicit analysis of the decisions facing people; (b) empirical assessment of individuals' relevant beliefs, values, and decision-making processes; and (c) development and empirical evaluation of communications focused on the facts critical to individuals' choices.
- ◆ Risk research should reflect the century of research into basic processes of judgement and decision making, both to ensure the robustness of its results and to take advantage of its knowledge for identifying and overcoming potential barriers to risk-related decisions.
- ◆ Unless measured appropriately, lay risk perceptions may be judged unfairly, leading professionals to be unduly critical of lay-people's decision-making capabilities.
- ◆ Only by understanding the decisions that individuals face can health research produce the information that people need.

Acknowledgement

The preparation of this chapter was supported by US National Science Foundation Grant SES 0433152. The views expressed are the author's.

References

- Bauman, K.E. (1980). *Predicting adolescent drug use: Utility structure and marijuana*. Praeger, New York.
- Beyth-Marom, R., Austin, L., Fischhoff, B. *et al.* (1993). Perceived consequences of risky behaviors. *Developmental Psychology*, **29**, 549–63.
- Black, W.C., Nease, R.F., and Tosteson, A.N.A. (1995). Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *Journal of the National Cancer Institute*, **8**, 720–31.
- Bostrom, A., Atman, C.J., Fischhoff, B. *et al.* (1994). Evaluating risk communications: Completing and correcting mental models of hazardous processes. Part 2. *Risk Analysis*, **14**, 789–98.
- Bostrom, A., Fischhoff, B., and Morgan, M.G. (1992). Characterizing mental models of hazardous processes: A methodology and an application to radon. *Journal of Social Issues*, **48**(4), 85–100.
- Brewer, N.T., Chapman, G.B., Gibbons, F.X. *et al.* (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, **26**, 136–45.
- Bruine de Bruin, W., Fischhoff, B., Halpern-Felsher, B. *et al.* (2000). Expressing epistemic uncertainty: It's a fifty-fifty chance. *Organizational Behavior and Human Decision Processes*, **81**, 115–31.
- Bruine de Bruin, W., Parker, A., and Fischhoff, B. (2007). Individual differences in adult decision-making competence (A-DMC). *Journal of Personality and Social Psychology*, **92**, 938–56.
- Budescu, D.F. and Wallsten, T.S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In *Decision making from the perspective of cognitive psychology* (eds. J.R. Busemeyer, R. Hastie, and D.L. Medin), pp. 275–316. Academic Press, New York.
- Byram, S., Fischhoff, B., Embrey, M. *et al.* (2001). Mental models of women with breast implants regarding local complications. *Behavioral Medicine*, **27**, 4–14.
- Christensen-Szalanski, J. and Bushyhead, J. (1993). Physicians' misunderstanding of medical findings. *Medical Decision Making*, **3**, 169–75.
- Dawes, R.M., Faust, D., and Meehl, P. (1989). Clinical versus actuarial judgment. *Science*, **243**, 1668–74.
- DeKay, M.L., Small, M.J., Fischbeck, P.S. *et al.* (2002). Risk-based decision analysis in support of precautionary policies. *Journal of Risk Research*, **5**, 391–417.
- Downs, J.S., Murray, P.J., Bruine de Bruin, W. *et al.* (2004a). An interactive video program to reduce adolescent females' STD risk: A randomized controlled trial. *Social Science and Medicine*, **59**, 1561–72.
- Downs, J.S., Bruine de Bruin, W., Murray, P.J. *et al.* (2004b). When "it only takes once" fails: Perceived infertility predicts condom use and STI acquisition. *Journal of Pediatric and Adolescent Gynecology*, **17**, 224.
- Eggers, S.L. and Fischhoff, B. (2004). Setting policies for consumer communications: A behavioral decision research approach. *Journal of Public Policy and Marketing*, **23**, 14–27.
- Erev, I. and Cohen, B.L. (1990). Verbal versus numerical probabilities: Efficiency, biases and the preference paradox. *Organizational Behavior and Human Decision Processes*, **45**, 1–18.
- Ericsson, K.A. and Simon, H.A. (1993). *Verbal reports as data*. MIT Press, Cambridge, MA.
- Feather, N. (1982). *Expectancy, incentive and action*. Erlbaum, Hillsdale, NJ.
- Fischhoff, B. (2005a). Cognitive processes in stated preference methods. In *Handbook of Environmental Economics* (eds. K-G Mäler and J. Vincent), pp. 937–68. Elsevier, Amsterdam.
- Fischhoff, B. (2005b). Decision research strategies. *Health Psychology*, **21**, S9–S16.
- Fischhoff, B. (1996). The real world: What good is it? *Organizational Behavior and Human Decision Processes*, **65**, 232–48.
- Fischhoff, B. (1995). Risk perception and communication unplugged: Twenty years of process. *Risk Analysis*, **15**, 137–45.
- Fischhoff, B. (1994). What forecasts (seem to) mean. *International Journal of Forecasting*, **10**, 387–403.
- Fischhoff, B. (1992). Giving advice: Decision theory perspectives on sexual assault. *American Psychologist*, **47**, 577–88.
- Fischhoff, B., Bruine de Bruin, W., Guvenc, U. *et al.* (2006). Analyzing disaster risks and plans: An avian flu example. *Journal of Risk and Uncertainty*, **33**, 133–51.
- Fischhoff, B., Downs, J., and Bruine de Bruin, W. (1998). Adolescent vulnerability: A framework for behavioral interventions. *Applied and Preventive Psychology*, **7**, 77–94.
- Fischhoff, B., Parker, A., Bruine de Bruin, W. *et al.* (2000). Teen expectations for significant life events. *Public Opinion Quarterly*, **64**, 189–205.
- Fischhoff, B. and MacGregor, D. (1983). Judged lethality: How much people seem to know depends upon how they are asked. *Risk Analysis*, **3**, 229–36.
- Fischhoff, B. and Quadrel, M.J. (1991). Adolescent alcohol decisions. *Alcohol Health & Research World*, **15**, 43–51.
- Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 552–64.
- Fischhoff, B., Slovic, P., Lichtenstein, S. *et al.* (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, **8**, 127–52.
- Fischhoff, B., Watson, S., and Hope, C. (1984). Defining risk. *Policy Sciences*, **17**, 123–39.

- Florig, K. and Fischhoff, B. (2007). Individuals' decisions affecting radiation exposure after a nuclear event. *Health Physics*, **92**, 475–83.
- Florig, H.K., Morgan, M.G., Morgan, K.M. *et al.* (2001). A deliberative method for ranking risks. *Risk Analysis*, **21**, 913–22.
- Fortney, J. (1988). Contraception: A life long perspective. In *Dying for love*. National Council for International Health, Washington, DC.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, **19**(4), 25–42.
- Gilovich, T., Griffin, D., and Kahneman, D. (eds.) (2003). *Judgment under uncertainty II: Extensions and applications*. Cambridge University Press, New York.
- Griffin, D., Gonzalez, R., and Varey, C. (2003). The heuristics and biases approach to judgment under uncertainty. *Blackwell handbook of social psychology*. Blackwell, Boston.
- HM Treasury. (2005). *Managing risks to the public*. Author, London.
- Howard, R.A. (1989). Knowledge maps. *Mananagement Science*, **35**, 903–22.
- Institute of Medicine (1986). *Confronting AIDS*. National Academy Press, Washington, DC.
- Kahneman, D., Slovic, P., and Tversky, A. (eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, New York.
- Krimsky, S. and Golding, D. (1992). *Theories of risk*. Praeger, New York.
- Koriat, A. (1993). How do we know that we know? *Psychological Review*, **100**, 609–39.
- Lerner, J.S. and Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, **81**, 146–59.
- Lichtenstein, S. and Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, **26**, 149–71.
- Lichtenstein, S., Fischhoff, B., and Phillips, L.D. (1982). Calibration of probabilities. In *Judgment under uncertainty: Heuristics and biases* (eds. D. Kahneman, P. Slovic, and A. Tversky), pp. 306–39. Cambridge University Press, New York.
- Lichtenstein, S., Slovic, P., Fischhoff, B. *et al.* (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 551–78.
- Linville, P.W., Fischer, G.W., and Fischhoff, B. (1993). AIDS risk perceptions and decision biases. In *The social psychology of HIV infection* (eds. J.B. Pryor and G.D. Reeder), pp. 5–38. Erlbaum, Hillsdale, NJ.
- Loewenstein, G., Weber, E., Hsee, C. *et al.* (2001). Risk as feelings. *Psychological Bulletin*, **127**, 267–86.
- McIntyre, S. and West, P. (1992). What does the phrase “safer sex” mean to you? Understanding among Glaswegian 18 year olds in 1990. *AIDS*, **7**, 121–6.
- Merton, R.F. (1987). The focussed interview and focus groups. *Public Opinion Quarterly*, **51**, 550–66.
- Merz, J., Fischhoff, B., Mazur, D.J. *et al.* (1993). Decision-analytic approach to developing standards of disclosure for medical informed consent. *Journal of Toxics and Liability*, **15**, 191–215.
- Morgan, M.G., Fischhoff, B., Bostrom, A. *et al.* (1992). Communicating risk to the public. *Environmental Science and Technology*, **26**, 2048–56.
- Morgan, M.G., Fischhoff, B., Bostrom, A. *et al.* (2001). *Risk communication: The mental models approach*. Cambridge University Press, New York.
- Murphy, A.H., Lichtenstein, S., Fischhoff, B. *et al.* (1980). Misinterpretations of precipitation probability forecasts. *Bulletin of the American Meteorological Society*, **61**, 695–701.
- National Research Council (2006). *Scientific review of the proposed risk assessment bulletin from the Office of Management and Budget*. National Academy Press, Washington, DC.
- National Research Council (1996). *Understanding risk: Informing decisions in a democratic society*. National Academy Press, Washington, DC.
- National Research Council (1989) *Improving risk communication*. National Academy Press, Washington, D.C.
- O' Hagan, A., Buck, C.E. Daneshkhah, A. *et al.* (2006). *Uncertain judgements: eliciting expert probabilities*. Wiley, Chichester.
- Peters, E. and McCaul, K.D. (eds.) (2005). Basic and applied decision making in cancer. *Health Psychology*, **24**(4), S3.
- Poulton, E.C. (1989). *Bias in quantifying judgment*. Lawrence Erlbaum, Hillsdale, NJ.
- Quadrel, M.J., Fischhoff, B., and Davis, W. (1993). Adolescent (in)vulnerability. *American Psychologist*, **48**, 102–16.
- Reimer, B., and Van Nevel, J.P. (eds). Cancer risk communication. *Journal of the National Cancer Institute Monographs*, **19**, 1–185.
- Reyna, V. and Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychology in the Public Interest*, **7**(1), 1–44.
- Riley, D.M., Fischhoff, B., Small, M. *et al.* (2001). Evaluating the effectiveness of risk-reduction strategies for consumer chemical products. *Risk Analysis*, **21**, 357–69.
- Ritov, I. and Baron, J. (1990). Satus quo and omission bias. Reluctance to vaccinate. *Journal of Behavioral Decision Making*, **3**, 263–77.
- Schwarz, N. (1999). Self reports. *American Psychologist*, **54**, 93–105.
- Shaklee, H. and Fischhoff, B. (1990). The psychology of contraceptive surprises: Judging the cumulative risk of contraceptive failure. *Journal of Applied Psychology*, **20**, 385–403.
- Slovic, P. (2001). *Perception of risk*. Earthspan, London.
- Slovic, P. and Fischhoff, B. (1983). Targeting risk. *Risk Analysis*, **2**, 231–8.
- Slovic, P., Fischhoff, B., and Lichtenstein, S. (1985). Characterizing perceived risk. In *Perilous progress: Managing the hazards of technology* (eds. R.W. Kates, C. Hohenemser, and J. Kaspersen), pp. 91–125. Westview, Boulder, CO.
- Slovic, P., Fischhoff, B., and Lichtenstein, S. (1979). Rating the risks. *Environment*, **21**(4), 14–20, 36–9.
- Slovic, P., Fischhoff, B., and Lichtenstein, S. (1978). Accident probabilities and seat-belt usage: A psychological perspective. *Accident Analysis and Prevention*, **10**, 281–5.
- Slovic, P., Lichtenstein, S., and Fischhoff, B. (1984). Modeling the societal impact of fatal accidents. *Management Science*, **30**, 464–74.
- Starr, C. (1969). Societal benefit versus technological risk. *Science*, **165**, 1232–8.
- Thaler, R. (1991). *Quasi-rational economics*. Russell Sage Foundation, New York.
- USEPA (1993). *A guidebook to comparing risks and setting environmental priorities*. Author, Washington, DC.
- Viscusi, K. (1992). *Smoking: Making the risky decision*. Oxford University Press, New York.
- Vlek, C. and Stallen, P.J. (1981). Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance*, **28**, 235–71.
- von Winterfeldt, D. and Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge University Press, New York.
- Willis, H.H., DeKay, M.L., Fischhoff, B. *et al.* (2005). Aggregate and disaggregate analyses of ecological risk perceptions. *Risk Analysis*, **25**, 405–28.
- Wogalter, M. (2006). *The handbook of warnings*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Woloshin, S., Schwartz, L.M., Byram, S. *et al.* (1998). Scales for assessing perceptions of event probability: A validation study. *Medical Decision Making*, **14**, 490–503.

