

Baruch Fischhoff

earthscan
from Routledge

RISK ANALYSIS AND HUMAN BEHAVIOR

The articles collected here are foundational contributions to incorporating behavioral research in risk analysis. On the one hand they address the questions of how to communicate with the public about risks, learning about the public's concerns and addressing them with responsive risk management and authoritative communications. On the other they address the questions of how to ensure that risk analyses make realistic assumptions about how people will behave, when they are part of risky systems, and what role expert judgment plays in analytical results. Ensuring behaviorally realistic risk analysis has been a central mission of the author's career. These articles touch on each fundamental area, demonstrating the possibilities for collaborative applied research and the basic research that it impels.

Baruch Fischhoff is Howard Heinz University Professor, Carnegie Mellon University, USA, Department of Social and Decision Sciences, Department of Engineering and Public Policy.

EARTHSCAN RISK IN SOCIETY SERIES

Series editor: Ragnar E. Löfstedt

- 1. Calculating Political Risk**
Catherine Althaus
- 2. The Citizens at Risk**
From Urban Sanitation to Sustainable Cities
Gordon McGranahan, Pedro Jacobi, Jacob Songsor, Charles Surjadi and Marianne Kjellen
- 3. The Earthscan Reader on Risk**
Edited by Ragnar E. Löfstedt and Åsa Boholm
- 4. The Ethics of Technological Risk**
Edited by Lotte Asveld and Sabine Roeser
- 5. Facility Siting**
Risk, Power and Identity in Land-Use Planning
Edited by Åsa Boholm and Ragnar E. Löfstedt
- 6. The Feeling of Risk**
New Perspectives on Risk Perception
Paul Slovic
- 7. Foresight in Action**
The Practice of Dealing with Uncertainty in Public Policy
Marjolein B. A. van Asselt, Susan A. van 't Klooster, Philip W. F. van Notten and Livia A. Smits
- 8. Global Environmental Risk**
Jeanne X. Kasperson and Roger E. Kasperson
- 9. Hazards, Vulnerability and Environmental Justice**
Susan L. Cutter
- 10. The Perception of Risk**
Paul Slovic
- 11. Public Safety and Risk Assessment**
Improving Decision Making
David J. Ball and Laurence Ball-King
- 12. Risk Governance**
Coping with Uncertainty in a Complex World
Ortwin Renn
- 13. Risk Management in Post-Trust Societies**
Ragnar E. Löfstedt
- 14. Risk, Media and Stigma**
Understanding Public Challenges to Modern Science and Technology
Edited by James Flynn, Paul Slovic and Howard Kunreuther
- 15. Risk, Uncertainty and Rational Action**
Carlo C. Jaeger, Ortwin Renn, Eugene A. Rosa and Thomas Webler
- 16. The Social Contours of Risk (Volumes 1 & 2)**
Jeanne X. Kasperson and Roger E. Kasperson
- 17. Social Trust and the Management of Risk**
Edited by George Cvetkovich and Ragnar E. Löfstedt
- 18. The Spatial Dimension of Risk**
How Geography Shapes the Emergence of Risks
Detlef Müller-Mahn
- 19. The Tolerability of Risk**
A New Framework for Risk Management
Edited by Frédéric Boudier, David Slavin and Ragnar E. Löfstedt
- 20. Transboundary Risk Governance**
Rolf Lidskog, Linda Soneryd and Ylva Uggla
- 21. Transboundary Risk Management**
Edited by Joanne Linnerooth-Bayer, Ragnar E. Löfstedt and Gunnar Sjöstedt
- 22. Trust in Cooperative Risk Management**
Uncertainty and Scepticism in the Public Mind
Michael Siegrist, Timothy C. Earle and Heinz Gutscher
- 23. Trust in Risk Management**
Uncertainty and Scepticism in the Public Mind
Edited by Michael Siegrist, Timothy C. Earle and Heinz Gutscher
- 24. Uncertainty and Risk**
Multidisciplinary Perspectives
Edited by Gabriele Bammer and Michael Smithson
- 25. Judgment and Decision Making**
Baruch Fischhoff
- 26. Risk Analysis and Human Behavior**
Baruch Fischhoff

RISK ANALYSIS AND HUMAN BEHAVIOR

Baruch Fischhoff



publishing for a sustainable future
LONDON AND NEW YORK

First published 2012

by Earthscan

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada

by Earthscan

711 Third Avenue, New York, NY 10017

Earthscan is an imprint of the Taylor & Francis Group, an informa business

© 2012 selection and editorial material, Baruch Fischhoff; individual chapters, Baruch Fischhoff and additional contributors

The right of Baruch Fischhoff to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

Fischhoff, Baruch, 1946-

Risk analysis and human behavior / Baruch Fischhoff.

p. cm. -- (Earthscan risk in society series)

Includes bibliographical references and index.

1. Risk perception. 2. Risk-taking (Psychology). 3. Decision making. 4.

Risk--Sociological aspects. I. Title.

HM1101.F57 2012

302'.12--dc23

2011027906

ISBN: 978-1-84971-442-6 (hbk)

ISBN: 978-1-84971-443-3 (pbk)

ISBN: 978-0-203-14071-0 (ebk)

Typeset in Bembo

by GreenGate Publishing Services, Tonbridge, Kent

CONTENTS

<i>List of figures</i>	<i>viii</i>
<i>List of tables</i>	<i>x</i>
<i>On the cover</i>	<i>xii</i>
<i>Introduction</i>	<i>xiii</i>

PART I

Overview 1

- | | |
|--|----|
| 1 Risk perception and communication | 3 |
| <i>Baruch Fischhoff</i> | |
| 2 Cognitive processes in stated preference methods | 33 |
| <i>Baruch Fischhoff</i> | |

PART II

Behaviorally realistic risk analysis 67

- | | |
|--|-----|
| 3 Cost–benefit analysis and the art of motorcycle maintenance | 69 |
| <i>Baruch Fischhoff</i> | |
| 4 Individuals' decisions affecting radiation exposure
after a nuclear explosion | 98 |
| <i>H. Keith Florig and Baruch Fischhoff</i> | |
| 5 Predicting emergency evacuation and sheltering behavior:
a structured analytical approach | 114 |
| <i>Matt Dombroski, Baruch Fischhoff, and Paul Fischbeck</i> | |

- 6 An integrative approach for label evaluation and design 140
Donna M. Riley, Baruch Fischhoff, and Mitchell Small
- 7 Analyzing disaster risks and plans: an avian flu example 149
Baruch Fischhoff, Wändi Bruine de Bruin, Ümit Güvenç, Denise Caruso, and Larry Brilliant

PART III

Social context of behavioral research 173

- 8 Lay foibles and expert fables in judgments about risk 175
Baruch Fischhoff, Paul Slovic, and Sarah Lichtenstein
- 9 Risk perception and communication unplugged: twenty years of process 208
Baruch Fischhoff
- 10 Non-persuasive communication about matters of greatest urgency: climate change 223
Baruch Fischhoff

PART IV

Risk communication 233

- 11 What forecasts (seem to) mean 235
Baruch Fischhoff
- 12 Parents' vaccination comprehension and decisions 260
Julie S. Downs, Wändi Bruine de Bruin, and Baruch Fischhoff
- 13 Development and evaluation of an HIV/AIDS knowledge measure for adolescents focusing on misconceptions 284
Wändi Bruine de Bruin, Julie S. Downs, Baruch Fischhoff, and Claire Palmgren
- 14 Communicating about xenotransplantation: models and scenarios 303
Wändi Bruine de Bruin, Ümit Güvenç, Baruch Fischhoff, Christopher M. Armstrong, and Denise Caruso

PART V	
Aiding individual risk decisions	321
15 What do patients want? Help in making effective choices— a patient's perspective <i>Baruch Fischhoff</i>	323
16 Giving advice: decision theory perspectives on sexual assault <i>Baruch Fischhoff</i>	327
17 Sticky decisions: peanut butter in a time of Salmonella <i>Gülbanu Kaptan and Baruch Fischhoff</i>	351
PART VI	
Aiding public risk decisions	361
18 A multi-channel stakeholder consultation process for transmission deregulation <i>Robin Gregory, Baruch Fischhoff, Sarah Thorne, and Gordon Butte</i>	363
19 The science and practice of risk ranking <i>Baruch Fischhoff and Granger Morgan</i>	379
20 Counting casualties: a framework for respectful, useful records <i>Baruch Fischhoff, Scott Atran, and Noam Fischhoff</i>	390
Postscript: integrating risk analysis and behavioral research	414
<i>Index</i>	416

FIGURES

1.1	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	7
1.2	Location of 30 hazards within the two-factor space obtained from League of Women Voters, student, active Club and expert groups	12
1.3	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	18
1.4	A simple decision tree for whether to ride with friends who have been drinking	20
1.5	A simple decision tree for whether to take saw palmetto for benign prostatic hyperplasia	21
2.1	Six “laws of the new psychophysics,” depicting the influence of experimental design	38
4.1	Factors affecting the utility of sheltering	105
4.2	Dose minimizing sheltering duration as a function of the dose-reduction factor of the sheltering structure and the time required to traverse the fallout-contaminated zone during evacuation	109
5.1	A schematic diagram of factors determining the outcomes of an RDD event	119
5.2	Academic expert estimates of compliance rates with an order to (a) shelter in place and (b) evacuate	121
5.3	Aggregated academic expert judgments of public behavior	123
5.4	Probability density functions for compliance with official orders to shelter in place or evacuate	127

7.1 Top-level risk model for pharmacological interventions and potential effects	156
7.2 Top-level risk model for a behavioral intervention	162
7.3 Second-level model for factors shaping compliance with a behavioral intervention	165
8.1 Current risk and benefit levels for selected technologies, as computed by Starr (1972)	179
8.2 Judgments of risk and benefit for activities and technologies used by Starr (1972)	179
8.3 Judgment of acceptable risk levels and perceived current benefit	180
8.4 Six laws of the new psychophysics	183
8.5 Relationship between judged frequency and the actual number of deaths per year for 41 causes of death	189
8.6 Judgments of perceived risk for experts and laypeople plotted against technical estimates of annual fatalities for 25 technologies and activities	195
8.7 Two dimensions underlying 18 qualitative aspects of risk	196
8.8 Qualitative aspects of risk loading on two factors	197
8.9 Validity of probabilistic precipitation forecasts	200
11.1 Upper and lower probability limits for probability expressions	237
11.2 Calibration curves for tasks varying in difficulty level	245
11.3 Confidence intervals surrounding estimated values of physical constants	246
11.4 Annual energy outlook forecasts	247
11.5 Confidence intervals for six production and consumption sectors	248
12.1 Integrated assessment of the effects of vaccination	264
12.2 Models implicit in communications about MMR vaccine	268
12.3 Models implicit in communications about the link between MMR vaccine and autism	270
12.4 Parents' beliefs about how vaccines work	272
13.1 Overconfidence in high-risk versus low-risk adolescents	295
14.1 Influence diagram showing recombination of a human and porcine virus	308
18.1 Expert model for consultation process	366
19.1 Some complexities of risk ranking, in a simple case	381
19.2 Location of 30 hazards within a two-factor risk space	385
19.3 Steps in the Q850 risk management decision-making process	388
20.1 Civilian deaths in Iraq, by month	400
20.2 Violent deaths recorded at Baghdad city morgue, by month	400
20.3 Israeli and Palestinian deaths from armed conflict, by quarter	402
20.4 Palestinians killed by Israelis, by quarter	403
20.5 Israelis killed by Palestinians, by quarter	404
20.6 Israeli and Palestinian deaths, by age	405

TABLES

1.1 Comparison of numerical verbal and statistical risk estimates	9
1.2 A standard multidimensional representation of risks	15
1.3 Data collection options for reader-based evaluations of risk communications	25
2.1 From artifact to main effect	36
2.2 Conditions favorable to respondents having articulated values	42
2.3 A general framework for characterizing environmental risks	52
2.4 A framework for defining transactions	54
2.5 Reasons for evaluating goods differently at different times	56
2.6 Manipulation checks	58
4.1 Estimated 10-year costs of provisioning a home shelter to sustain four people with three days of supplies	101
4.2 Event string required for home shelter preparation to be useful	103
4.3 Travel recommendations for the period immediately after detonation of a 10-kiloton fission bomb, before the arrival of fallout	107
5.1 Hurricane evacuation rates for several major hurricanes hitting the United States	116
5.2 Best-guess compliance rates for eight scenarios, elicited from emergency coordinators	126
6.1 Selected health effects and precautionary behaviors mentioned on paint-stripper labels	144
6.2 Location and ventilation strategies for reading scenarios	146
6.3 Peak exposure (ppm)	147
7.1 Excerpts from an (unstructured) avian flu scenario	155
7.2 Questions to ask in checking a model's clarity	158
7.3 A model-based scenario	164
8.1 Ways that an elicitor may affect a respondent's judgments of value	186
8.2 Two formulations of a choice problem	187

8.3 Bias in perceived frequency	191
8.4 Some common problems leading to underestimation of risks	202
9.1 Developmental stages in risk management (ontogeny recapitulates phylogeny)	209
11.1 Interpretations of AIDS risk questions	239
11.2 Intuitive interpretations of ambiguous event descriptions	240
11.3 The pedigree of numerical estimates	251
11.4 Where did the Weather Channel go wrong?	252
11.5 Risks of (mis)communication	254
12.1 Sample questions from interview protocol	266
13.1 Domains of interview protocol and sample questions	287
13.2 Demographic information and sexual behaviors of interviewees	288
13.3 Example quotes describing misunderstood concepts	290
13.4 Demographic information and sexual behaviors of knowledge test respondents	294
14.1 One of the three scenarios	309
14.2 Mean (<i>SD</i>) for dependent measures	312
19.1 Risk comparisons	383
19.2 A standard risk characterization	386

ON THE COVER

The 'Perception of Risk' series borrows from the science of Risk Perception by using an x-y axis to map out and visualize perceptions of risk to myself and to others, real and/or imagined. Each charting varies according to the date and place it was made.

Adele Henderson

Perception of Risk, Buffalo, New York, 7/10/09, 2010

Letterpress (2 colors; hand-set type) and lithography (7 colors, 6 plates and 1 stone), hand printed on Rives BFK, 14" × 18"

www.adelehenderson.com

INTRODUCTION

I was drawn to studying risks because the problems are so messy. They require understanding both people and problems. They require understanding questions of both fact and values. They require grappling with uncertainties about both the world (what can happen) and oneself (what do I want). They require both learning from others and challenging their views. They require both critical analysis (not taking claims at face value) and constructive synthesis (not being paralyzed by complications). They require mastery of both form (general decision rules) and substance (specific risky processes). Even when there are formulae for calculating some aspects of a risk decision, there is nothing formulaic about their application. Seemingly straightforward tasks (e.g., counting casualties) often raise tough questions (e.g., whom to include in the count, how to tell how complete it is), which can take different forms for different risks.

However, because decisions must eventually be made, the deliberations over these thorny issues cannot be interminable. At some point, one must weigh the evidence and reach some conclusions, recognizing the stakes riding on them. For a scientist, those stakes not only give the work added meaning, but can lead to uncommon, rewarding collaborations with other scientists and practitioners. Basic research naturally fosters rivalries, both within and across disciplines, over which problems matter most and which methods are best for addressing them. Risk decisions can subordinate those rivalries to the practical mission of assembling the most relevant expertise, wherever it lies.

As a result, studying risk decisions means working with diverse professionals, trying to solve intellectually challenging problems, and learning the fascinating details of some risk domain, in order to help real people. Moreover, the work can offer unique opportunities to advance the basic sciences involved. As noted by Alan Baddeley (1979), former head of the Medical Research Council

Applied Psychology Unit, healthy sciences require *applied basic* research, examining how well their theories work in practice, and *basic applied* research, answering fundamental questions raised by practical problems (Baddeley, 1979). That tension between the general and the particular is what keeps risk research from being “just” consulting, solving “real world” problems, but without stepping back to think and generalize.

When the opportunity arose to create a collection of my work on risk, I sought readings that captured some of the excitement (and occasional frustration) of trying to serve worlds of theory and practice. Each reading includes both applied contexts and attempts to produce general knowledge. Doing the research that the readings report has allowed me to collaborate with some remarkable individuals, committed to “getting it right” (to use Amos Tversky’s phrase) in the academic world, the “real” world, or both. I hope that readers get similar chances.

PART I

Overview

Risk decisions involve choices among options, each with some chance of producing some amount of some outcomes that decision makers value. Decision makers' beliefs about those outcomes are commonly called their *risk perceptions*. The relative weight that they assign to them is captured in their *risk evaluations*. Scientific assessment of the outcomes' probability and value is called *risk analysis*. Sharing information about them is *risk communication*.

In these terms of art, “risk” is something of a misnomer, implying that a decision could be about risks alone. There is, however, no reason to choose a risky option unless it offers some possible benefit, if only the chance to reduce a larger risk. For example, one might accept the risk of coronary bypass surgery in the hope of having less risk of stroke. Although a risk decision must have at least one unpredictable outcome (e.g., dying in surgery), it may also have certain outcomes, both good ones (e.g., knowing that one has done everything possible) and bad ones (e.g., cost, discomfort). Thus, “risk decisions” should be read as “decisions potentially having certain and uncertain positive and negative outcomes.”

The two readings in this section offer overviews of research into how people perceive and evaluate risks. That research not only describes behavior, but compares it to *normative* standards for how such judgments should be made. Having such standards shows where people need help and poses the question of why they need it. (Do they lack some fundamental ability? Have their lives failed to teach them better ways to make decisions? Do social pressures lead people to act against their better judgment?) Creating those standards requires the kind of formal analysis that is part of any full decision science inquiry. Employing the standards may entail either comparisons with an external norm (offering a right answer) or tests of internal consistency.

With risk perceptions (Chapter 1), an archetypal external test might be comparing probability of precipitation forecasts with how often it rains. (For experienced weather forecasters the two are usually quite close.) A corresponding test of internal consistency might compare forecasts of the probability of rain on one day and on that day or the next (which should be at least as large). Applying such tests involves the kind of combined empirical and analytical research that distinguishes decision science approaches to risk. It can guide risk communications designed to improve flawed risk perceptions or to reinforce appropriate ones. That may mean providing better risk estimates, for people who just need a number (e.g., lottery odds, drug side effect rates). Or, it may mean improving their *mental models* of the processes creating and controlling risks—so that the estimates make intuitive sense.

With risk evaluations (Chapter 2), typical external tests compare what people value to socially defined norms regarding what should matter to them (e.g., they should care about their health, save for retirement, or respect their elders). Typical internal comparisons are between willingness to pay for a small amount of a good and a larger amount (which should be at least as big). Here, too, decision makers may just need help in the form of a summary (e.g., most people prefer annuities, our faith proscribes that food). Or, they may need the deeper understanding involved in *constructing* their preferences, figuring out what they want in a specific choice (e.g., how much they will pay for an auto safety feature), based on their general values (e.g., wanting safety and money).

Chapter 1 focuses on health decisions and Chapter 2 on environmental decisions. However, the basic issues are similar in the two domains—and in many others. As diverse as risks may be, decision science addresses them with a common set of analytical and empirical methods.

1

RISK PERCEPTION AND COMMUNICATION

Baruch Fischhoff

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 laypeople 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 transfats, 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 (Loewenstein *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 and 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 – as 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 1.1 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.
- (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 1,000 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,

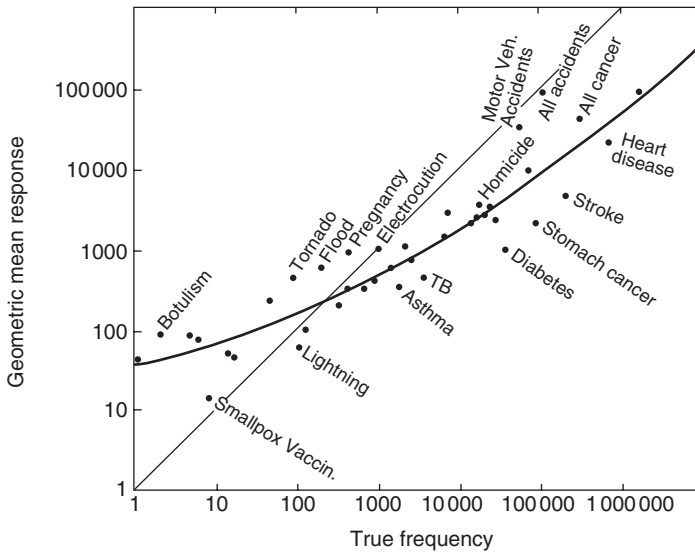


Figure 1.1 Best quadratic fit line to geometric mean judgements of the annual toll from 40 causes of death in the United States, compared to best available statistical estimates.

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

individuals' estimates typically ranged over 3–4. That compression could reflect anchoring, if judgements were drawn towards 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, probably 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. Koriati 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, 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 and Wallsten 1995; Schwarz 1999).

Table 1.1 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 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 and 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.

Table 1.1 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)

- (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 health-care 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 and 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 such as 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 1.2 plots factor scores within the common factor space for these four groups.

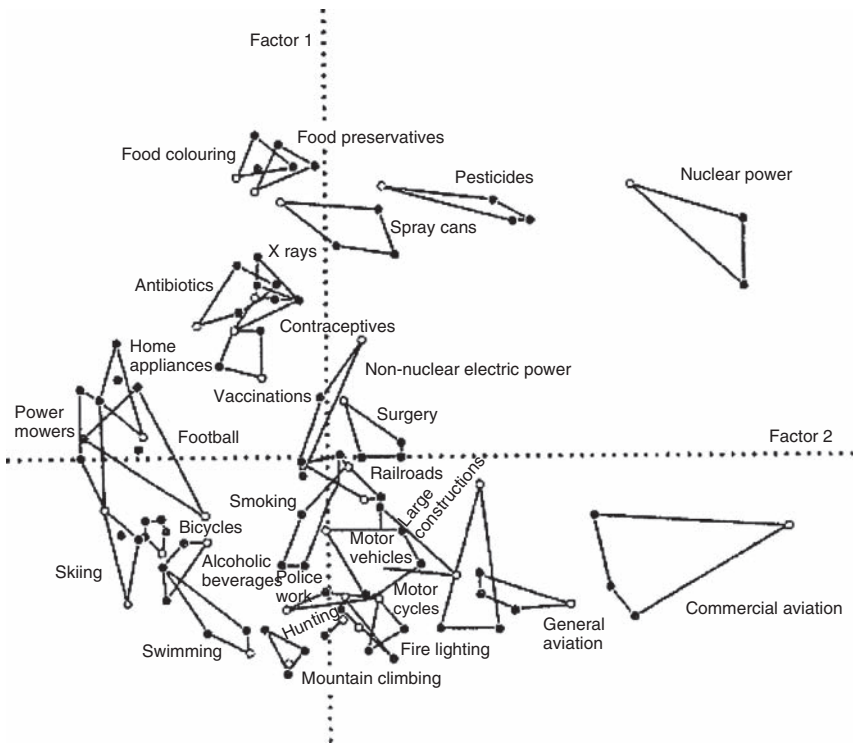


Figure 1.2 Location of 30 hazards within the two-factor space obtained from League of Women Voters, student, active Club and expert groups. Respondents evaluated each activity or technology on each of nine features. Ratings were subjected to principal components factor analysis, with a varimax rotation. 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).

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 towards 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 and 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 table-spoonful of peanut butter and 50 years living by a nuclear power plant create a one-in-a-million risk of premature death). Box 1.1 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 1.2). The UK government has endorsed a variant (HM Treasury 2005).

Box 1.1

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 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 1.2 A standard multidimensional representation of risks

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

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 and 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 and 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 1.3 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, 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

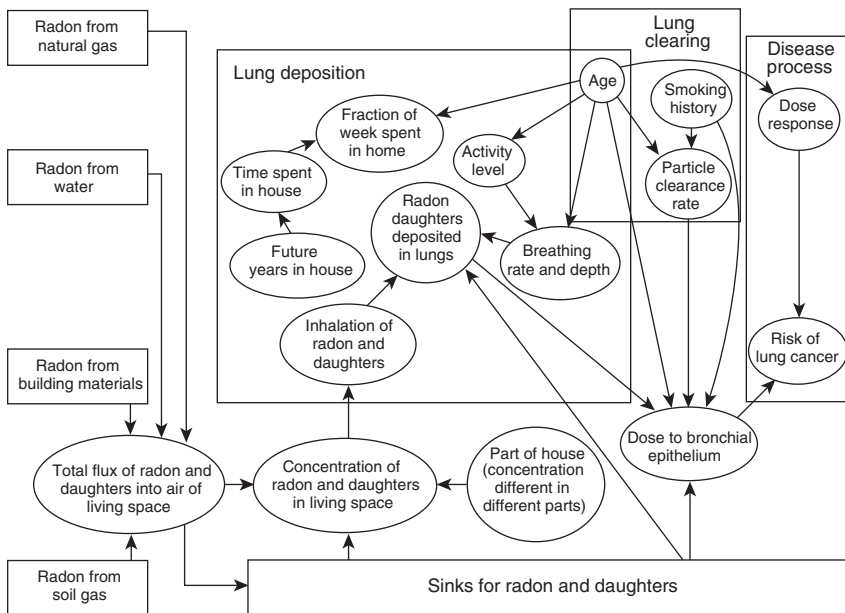


Figure 1.3 Expert influence diagram for health effects of radon (in a home with a crawl space). This diagram was used as a standard and as an organizing device to characterize the content of lay mental models.

Source: Morgan *et al.* (1992).

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. 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 (1982) 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 1.4 and 1.5 show 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 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.

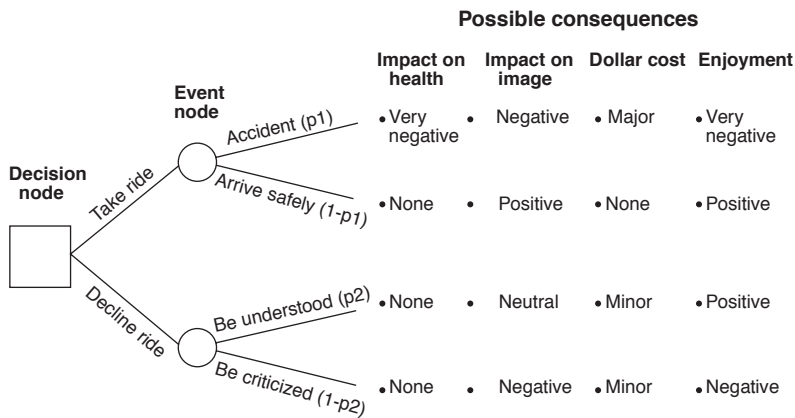


Figure 1.4 A simple decision tree for whether to ride with friends who have been drinking.
Source: Fischhoff and Quadrel (1991).

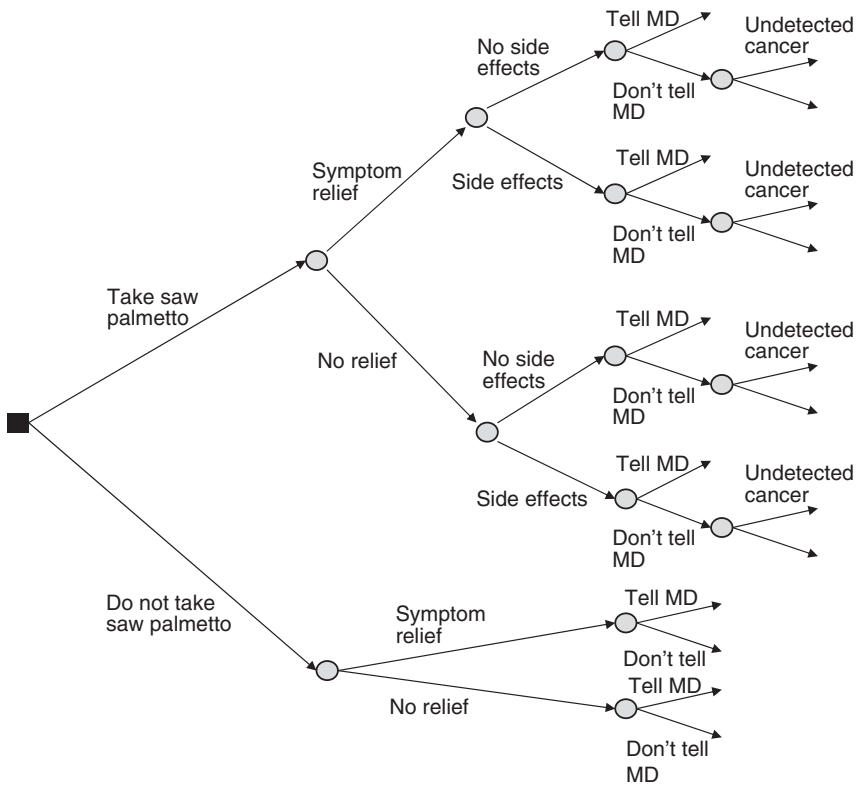


Figure 1.5 A simple decision tree for whether to take saw palmetto for benign prostatic hyperplasia.

Source: Eggers and Fischhoff (2004).

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 Figure 1.4. 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 and Baron 1990).

Different methods have different strengths and weaknesses (Ericsson and 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, 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 earlier: (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 four 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 can 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 1.3 summarizes approaches to reader-based evaluation. Open-ended interviews are the best way to reduce these threats. However, performing them to scientific standards is labour intensive. It entails conducting, transcribing and coding interviews, with suitable reliability checks, in addition to producing the expert model needed for their evaluation.

Table 1.3 Data collection options for reader-based evaluations of risk communications

	<i>Strengths</i>	<i>Weaknesses</i>
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)

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 feed back 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 their pet theories of how the public needs to be alarmed or calmed. Without qualified experts, these roles will be filled by amateurs, imperilling 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, and new research creating the basis for risk analyses.

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., 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.
- 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.
- 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 J.R. Busemeyer, R. Hastie and D.L. Medin (Eds.), *Decision Making from the Perspective of Cognitive Psychology*. New York: Academic Press, pp. 275–316.
- 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., Bruine de Bruin, W., and Fischhoff, B. (2008). Patients' vaccination comprehension and decisions. *Vaccine*, 26, 1595–1607.
- 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.
- 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.
- 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*. Cambridge, MA: MIT Press.