

The International
Library of Ethics,
Law and Technology 5

Sabine Roeser
Editor

Emotions and Risky Technologies

 Springer

Emotions and Risky Technologies

The International Library of Ethics, Law and Technology

VOLUME 5

Editors

Anthony Mark Cutter, *Centre for Professional Ethics, University of Central Lancashire, United Kingdom*

Bert Gordijn, *Ethics Institute, Dublin City University, Ireland*

Gary E. Marchant, *Executive Director, Center for the Study of Law, Science, & Technology, University of Arizona, USA*

Alain Poupidou, *Former President, European Patent Office, Munich, Germany*

Editorial Board

Dieter Bimbacher, *Professor, Institute of Philosophy, Heinrich-Heine-Universität, Germany*

Roger Brownsword, *Professor of Law, King's College London, UK*

Ruth Chadwick, *Director, ESRC Centre for Economic & Social Aspects of Genomics, Cardiff, UK*

Paul Stephen Dempsey, *Professor & Director of the Institute of Air & Space Law, Université de Montréal, Canada*

Michael Froomkin, *Professor, University of Miami Law School, Florida, USA*

Serge Gutwirth, *Professor of Human Rights, Comparative Law, Legal theory and Methodology, Faculty of Law, Vrije Universiteit, Brussels, Belgium*

Henk ten Have, *Director, UNESCO Division of Ethics of Science and Technology, Paris, France*

Søren Holm, *Director, Cardiff Centre for Ethics, Law & Society, Cardiff, UK*

George Khushf, *Humanities Director, Center for Bioethics, University of South Carolina, USA*

Justice Michael Kirby, *High Court of Australia, Canberra, Australia*

Bartha Maria Knoppers, *Chair in Law and Medicine, Université de Montréal, Canada*

David Krieger, *President, The Waging Peace Foundation, California, USA*

Graeme Laurie, *Co-Director, AHRC Centre for Intellectual Property and Technology Law, UK*

Rene Oosterlinck, *Director of External Relations, European Space Agency, Paris*

Edmund Pellegrino, *Chair, President's Council on Bioethics, Washington, DC, USA*

John Weckert, *Professor, School of Information Studies, Charles Sturt University, Australia*

For further volumes:

<http://www.springer.com/series/7761>

Sabine Roeser
Editor

Emotions and Risky Technologies

 Springer

Editor

Sabine Roeser
Department of Philosophy
Delft University of Technology
2628 BX Delft
Netherlands
s.roeser@tudelft.nl

ISSN 1875-0044 e-ISSN 1875-0036
ISBN 978-90-481-8646-4 e-ISBN 978-90-481-8647-1
DOI 10.1007/978-90-481-8647-1
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010924810

© Springer Science+Business Media B.V. 2010

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

In Memory of Robert Solomon

Foreword

“Acceptable Risk” – On the Rationality (and Irrationality) of Emotional Evaluations of Risk

What is “acceptable risk”? That question is appropriate in a number of different contexts, political, social, ethical, and scientific. Thus the question might be whether the voting public will support a risky proposal or project, whether people will buy or accept a risky product, whether it is morally permissible to pursue this or that potentially harmful venture, or whether it is wise or prudent to test or try out some possibly dangerous hypothesis or product. But complicating all of these queries, the “sand in the machinery” of rational decision-making, are the emotions.

It is often noted (but too rarely studied) that voters are swayed by their passions at least as much as they are convinced by rational arguments. And it is obvious to advertisers and retailers that people are seduced by all sorts of appeals to their vanities, their fears, their extravagant hopes, their insecurities. At least one major thread of ethical discourse, the one following Kant, minimizes the importance of the emotions (“the inclinations”) in favor of an emphatically rational decision-making process, and it is worth mulling over the fact that many of those who do not accept Kant’s ethical views more or less applaud his rejection of the “moral sentiment theory” of the time, promoted by such luminary philosophers as David Hume and Adam Smith. (Jean-Jacques Rousseau should also be included here, but Rousseau’s dim view of science and technological progress would rather complicate the discussion.) Sentiments such as “sympathy” are too undependable, the critics say, to provide a secure basis for important moral decisions.

Science, finally, even shorn of its moral implications and the delicate questions of public support, has ample reason to worry about the consequences of seriously entertaining hypotheses that impact future scientific research (not to mention public health and more general questions of welfare). Questions surrounding global warming, for instance, involve terrifying risks, but one could (if one wished) describe these in solely scientific terms regarding the fate of the earth and such purely technical matters. But enough tongue-in-cheek reductionism: science is *not* separable from morals and values more generally, and it is not free of emotional complicity. That one engages in science at all (whether because of “scientific curiosity,” required

courses, or professional ambition) requires motivation and interest, and that one pursues a particular field or a particular problem or puzzle requires focused inspiration and drive. Hypotheses do not merely spring from existing technology and scientific theory. They express interests, and these interests are often the convergence of any number of emotions. Global warming is not just a scientific curiosity. It is something that makes any sensible person extremely anxious, and it shames (or ought to shame) all of us who are such extravagant degraders of the environment. (How much CO₂ did you burn getting to your last academic conference?)

That last parenthetical example, however, introduces the point that I would like to pursue here in this foreword. And that is that the complexities introduced by emotional involvement in our consideration of risky technologies should not be considered as distortions or interferences in our thinking (“sand in the machinery”) but as an essential part of our rational reflections. The question is *which* emotion or emotions are engaged and *how*. Shame is not in itself a “negative emotion.” As Aristotle pointed out a long time ago (before risky technologies became a major ethical concern), the capacity for shame is an essential part of being a good person. To be motivated by shame to correct either one’s behavior or the consequences of one’s behavior does not discredit one’s intentions but rather ennobles them. To pursue and encourage the science of global warming because we are and ought to be ashamed of our contributions to the problem is what it means to be a good global citizen these days.

I start with the example of global warming to make a general point about the role of emotions in science and more generally in evaluating risk. The question is not whether our decisions are emotional or not but rather how our emotions shape our evaluations of risk. Where technology enters into the discussion is worth noting. There are all sorts of risks in the world (climbing a volcano on the brink of eruption, invading a country on the brink of civil war, lying to your spouse about where you were last weekend, eating uncooked meats or seafood), but risky technology occupies a curious place in our understanding of dangers and probabilities. This is because risky technology is both risky (outcomes and consequences are uncertain) and within our control (the root of the term is *techné*, craft or skill). It is not just a question about how we should respond to a risk (should we go ahead or not, should we try it or not?) but what we do to *create* the risk. In other words, technological risk is both a matter of choice, control, and responsibility, and uncertainty and insecurity. Thus among the risky technologies most often cited are the genetic technologies of cloning, medical procedures using unproved drugs or surgical procedures, genetically engineered crops and new antibiotics, and untested environmental “solutions” to ecological problems. All of these present us with not only choices and risks but provoke powerful emotions that may or may not be rational.

I do not know much about risk management (although, like all more or less rational creatures, I *practice* it all the time). Nor do I know much more than the average magazine and website reader about the new technologies as such. But what I do know something about and have thought about a great deal is the rationality of emotions. And since risk assessment is both a matter of rationality and of emotion that is what I would like to focus on here.

The Rationality of Emotions

Traditionally, emotions have not been thought of in terms of rationality. On the contrary, emotions have been thought of as paradigms of irrationality, which is why they have so often been thought to be interferences to rational deliberation and contrasted with reason. Thus it is all too common to hear “well, intellectually [rationally] I know that, but emotionally I just can’t bring myself to believe it.” Thus the familiar ironies of ordinary risk assessment: some people are terrified at the prospect of flying in a commercial airliner but know full well (and in fact may know in much more detail than most people) the encouraging statistics regarding the risk and likelihood of an accident. (It is well-known and often said that the odds make it much more likely that one will have a fatal accident on the way to the airport than die in a plane crash.) So, too, regarding fatal bee stings, being struck by lightning, spiders, some suspected carcinogens, signs of aging, nuclear war and terrorism. Calculating the odds of such occurrences is a controversial procedure in itself, but estimating the emotional weight of the possibilities is quite another matter. As the “fear of flying” example shows all so clearly, not only are the “objective” risks quite at odds with the emotional [“subjective”], but it is wholly possible that one can be fully aware of the former without measurable effect on the latter.

The usual way of treating such divergence (a “paradox” to theorists for whom self-deception and incontinence present logical difficulties) is to insist on the divorce between intellect and emotion (“rationally I know that, but emotionally. . .”). I have argued against this at considerable length (in *True to Our Feelings*) on the grounds that (a) such examples are anomalous and not typical, much less paradigms supporting a general opposition between rationality and emotion and (b) I think that the opposition in such cases should not be understood in terms of the distinction between reason and emotion but rather in terms of two levels of belief or appraisal, a largely unconscious or “embodied” appraisal which may be to a certain extent “hard-wired” and based in the “lower” (sub-cortical) parts of the brain and a fully conscious and more or less deliberative evaluation that brings in evidence, rational argument, and putatively “objective” risk assessment. But the more important point is that rationality is not opposed to emotions but always in conjunction with them. That is, there are certain things that it is *right* to feel, both correct and warranted by the circumstances. There are *good reasons* for being fearful, and good reasons for being hopeful as well as good reasons for feeling responsible, regretful, grateful, ashamed, angry and so on. Thus one might distinguish between two different kinds of reasons (actually, an entire spectrum of reasons), “gut level” reasons on the one hand (“truthiness” on one current theory) and deliberative reasons, reasons that mention evidence, use inferences, and tend to be more or less fully articulate.

Emotions can be said to be rational not only in the sense that they are correct and warranted but also in the sense that they are functional. It has become something of a platitude in contemporary psychology that our emotions or at any rate the capacity to have emotions evolved and satisfied the conditions of natural selection. It does not follow that emotions (much less particular emotions) were distinctively “selected for” or that they still serve the purposes which may have once made them valuable

in the past (Darwin's "once serviceable habits"). For instance, anger may have been a valuable emotion by way of stoking one's aggression in prehistoric circumstances but it may be deleterious or generally dysfunctional in a modern urban environment. And emotions (and particular emotions) may well be bi-products of other evolved traits. Nevertheless, as a general rule, emotions are functional. They serve a purpose and play an important role in our personal and social lives. To be sure, emotions may sometimes be disruptive and harmful, but they are also motivating and useful. Thus Hume insists that only our passions, and not reason alone, can motivate us to be moral, or, for that matter, to do anything. Modern neurology, for instance, in the work of Antonio Damasio, comes to much the same conclusion.

Accordingly, emotions can be rational in the sense that they serve our ends, not only fortuitously or by way of evolution but because they can be used as means to achieve our ends in life. We can cultivate our emotions and our emotional reactions by cultivating our "character", and character in turn depends on what emotions one cultivates, how he or she orients himself/herself in the world. Emotions do not just "happen" to us. They are the product of not only our culture but of our behavior and our attitudes over time. We are thus, to a certain extent, responsible for our emotions. They are valuable not only in an evolutionary sense, as what Oatley and Jenkins call "ready repertoires of action", but in terms of our individual projects and aspirations as well. Getting angry may be an important step in motivating oneself to face obstacles and overcome them. Falling in love is an important step in forming an intimate relationship. Emotions provide both the substance of a good life and its directives. Many psychologists rightly worry about how we "cope" with our emotions, but it is equally important to appreciate how we use them. Jean-Paul Sartre argues that emotions are strategies. They do not just happen to us but we use them to manipulate others and, more importantly, to maneuver ourselves into ways of thinking and acting that suit our goals and self-image.

Emotions are or can be rational not only in the sense that they can be correct in their identification of their object, insofar as they are appropriate to the circumstances and appropriate in the present cultural setting, and insofar as they are fair and warranted. They also are or can be rational insofar as they are suitable to furthering our goals and fitting with our self-image. Getting angry at one's boss may be thoroughly warranted but still irrational insofar as it frustrates one's career goals. A Buddhist monk may be fully justified in getting jealous of a fellow monk but his jealousy is nevertheless irrational insofar as it is incompatible with his conception of himself as a Buddhist. Aristotle, accordingly, urges us to cultivate our characters and insists that being virtuous is having the right emotions, at the right time, in the right circumstances, and to the right degree. Having the right emotions is thus a key ingredient in living well.

The role of emotions in risk assessment, accordingly, is much more complex than the "reason vs. emotion" paradigm would suggest or can appreciate. In the above examples of fear—fear of flying, fatal bee stings, being struck by lightning, spiders, some suspected carcinogens, signs of aging, nuclear war and terrorism—there is a remarkable range of reasons involved, some of which are "gut reactions" but most of which involve some education as to the specifics of the situation. For some people,

the fear of fatal bee stings is perfectly reasonable. (For most it is not.) The fear of being struck by lightning can become obsessional, to be sure, but for the most part it is a good reason to take certain precautions. But technological risk, when it is not just something to be worried about but a project in which one participates (even if just by voting for it or giving one's support in some other way), involves a quite different range of reasons, namely reasons for *doing something*, reasons involving responsibility and choice.

Emotions and Responsibility

The concept of responsibility is not usually considered an issue relevant to the emotions, although, needless to say, one's responsibilities (and their fulfilment or failure) may well become the *object* of such emotions as anxiety, guilt, pride, shame, regret, and so on. But it would be a mistake to overly separate the emotion and the object, and the concept of responsibility can also be an ingredient in the emotion, what gives the emotion its shape and determines its object. To repeat a point I made earlier, risky technology is both risky (outcomes and consequences are uncertain) and within our control (*techné* as craft or skill). Thus the emotions concerning risky technology involve choice, control, and responsibility as well as uncertainty and insecurity. They involve the question *what to do*. The emotions in risk assessment, accordingly, are much more complex than any emotions that might be classified as mere *reactions*. They are anticipatory, and to some extent, no matter how hopeful they may be, they will involve a certain amount of *anxiety*. (I here follow Jean-Paul Sartre in distinguishing anxiety from fear in that the latter involves a vulnerability to the world whereas the former involves uncertainty as what one is to *do*.) Thus the awareness of one's own responsibility (again whether it is in the actual design and implementation of a risky technology or in the decision whether or not to support it) is essential to one's emotional attitudes toward risky technologies.

How does one's sense of responsibility enter into various emotions, especially the emotions associated with risk? What must be assumed, in any such analysis, is that self-consciousness must be an essential part of any such emotion, and self-consciousness in a fairly sophisticated sense. It has been argued (for example, by Damasio in his *The Feeling of What Happens*) that *some* sense of self reaches far down the phylogenetic ladder. But the sense of self-consciousness that is required in order to have a sense of responsibility, or an emotion of responsibility, involves much more than evidence of organic self-organization (as in Damasio). Many philosophers would argue that it requires language, and a language of responsibility in particular, a language that includes such moral concepts as "blame" and "accountability." Whether or not one goes quite this far, I think that the arguments for the sophistication of such emotions as guilt, shame, remorse, and pride are substantial.

Embarrassment is an odd case, since it seems to imply the *lack* of responsibility. That could be pushed either way. One might emphasize those cases in which one is well aware that one is not responsible for one's situation, or one might emphasize

those cases in which there is no thought of responsibility. In the latter case, it might make good sense to say that dogs and cats might feel embarrassment, say, after a humiliating incident, but not in the former. Dogs and [especially] cats do not have a sense of responsibility. Shame and guilt provide more interesting cases. Dogs seem to feel shame and guilt when they have done something wrong, but shame seems to involve a much more robust sense of self than dogs are capable of, and their seeming guilt has often been interpreted, quite convincingly, as anticipation of punishment.

But to get back to the topic at hand, such moral emotions that presuppose the concept of responsibility are the key to questions about emotions and risky technologies. This has important implications for the analysis of moral emotions in general. It is not as if moral evaluation is built on top of a perception of the fact (or in this case, the possibilities) which in turn provokes the appropriate emotions. Our emotional responses regarding risky technologies are a holistic amalgam of perception of possibilities and moral evaluation. And these are not “components” or separable ingredients but all of a piece. In particular, judgments of responsibility (what might happen and who would be responsible) are an intrinsic and inseparable aspect of such emotions. Not that one could not in some sense make “the same” judgments in a wholly disinterested and dispassionate way, that is, affirm the same propositions based, perhaps, on a cold calculation of probabilities. But if it is the emotion that interests us, the judgment would not be “the same” at all, but rather the dissociated conclusion of a process that is not at all engaged in the situation. (There is always the possibility of pathological dissociation with repressed emotional involvement, but I do not want to consider this possibility here.)

So what are the appropriate emotions with regard to risky technologies? It depends, of course, on the specific technology and its circumstances. But my point here is that the appropriate emotions will in any case include a serious consideration of one’s responsibilities and not just the fear of consequences or the uncertainty that is always associated with risk. Fear is, to be sure, an important emotion when potential catastrophe is a real life possibility. But anxiety is not the same as fear, and the anxiety that is always appropriate in the consideration of risky technologies is the anxiety that comes with the responsibility of bringing such a technology into being or using it in more or less novel circumstances. It is a quite distinctive dimension of some emotions – this sense of responsibility, and to omit it from our analysis would suggest that all of us are nothing more than victims of risky technologies rather than their willing authors, supporters, and consumers. That is what Sartre rightly called “bad faith,” and it represents a total abnegation of responsibility regarding those circumstances that these days define a good part of our lives, and may well end our lives as well.

Austin, Texas

Robert C. Solomon

Note by Sabine Roeser

With thanks to Robert Solomon’s wife, Kathleen Higgins, for allowing me to publish his essay for the conference on *Moral Emotions about Risky Technologies*

(May 3–4, 2007) as a foreword to this volume, which grew out of the same conference. Unfortunately, Robert Solomon was not able to share his thoughts about this topic with us at the conference. He passed away on January 2nd 2007. But his ideas about the rationality of emotions remain with us. This volume focuses on emotions and their (ir)rationality, in relation to risky technologies.

Acknowledgements

This volume grew out of a conference on “Moral Emotions about Risky Technologies” that I organized at the Philosophy Department of Delft University of Technology on May 3 and 4, 2007. The conference was generously sponsored by the KNAW (Royal Dutch Academy of the Sciences), the Platform for Ethics and Technology of TU Delft and the Philosophy Department of TU Delft. My work for editing this volume and writing the introduction and my chapter for this volume has been sponsored by the Netherlands Organization for Scientific Research (NWO), with VENI-grant number 275-20-007.

I am very grateful to the contributors to the conference and to this volume who were willing to think about this new topic. It has been very inspiring to listen to and read about their ideas. Thanks to *Judgment and Decision Making* for allowing to reprint the paper by Paul Slovic and to *Pennsylvania Law Review* for allowing to reprint the paper by Dan Kahan. I would like to thank Peter Kroes and Jeroen van den Hoven of the Philosophy Department at TU Delft for the ongoing support they provide for my work. They create a truly ideal work environment. Once again, Henneke Piekhaar did an incredible job in supporting me with the organization of the conference. I would like to thank Marco Eliens for his outstanding work as an editorial assistant during the preparation of this volume. I would like to thank the Series editors, especially Bert Gordijn, and the various staff members at Springer and Anandhi Bashyam for their excellent support in the completion of this volume.

Personal acknowledgements might fit better with a monograph than with an edited volume. Nevertheless, I want to thank my wonderful family for the warmth with which they surround me every moment of my life. Parker and Mae were born during the period that I worked on my VENI-research-project of which this volume is the final result. Jeff, you have been with me for so many years now. Thank you for your love and intellectual kinship. You three are the evidence of the importance of emotions.

Robert Solomon has been the philosopher who, with intellectual passion, has put the role of emotions highly on the philosophical agenda. I was extremely honored that he was willing to contribute to my conference and a possible edited volume.

Unfortunately, he passed away a few months before the conference, on January 2nd 2007. His death is an incredible loss for the community of emotion-researchers. This volume is dedicated to his memory.

Delft, The Netherlands
June 2009

Sabine Roeser

Contents

Emotions and Risky Technologies: Introduction and Overview	xxi
Part I Emotions as Distortions About Risk	
Moral Heuristics and Risk	3
Cass R. Sunstein	
Here’s How I Feel: Don’t Trust Your Feelings!	17
Ronald de Sousa	
If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide	37
Paul Slovic	
Marketing Risk: Emotional Appeals Can Promote the Mindless Acceptance of Risk	61
Ross Buck and Whitney A. Davis	
Emotions as Aids and Obstacles in Thinking About Risky Technologies	81
Dylan Evans	
Part II Emotions and Virtues in Risk Assessment	
Risk Assessment as Virtue	91
Sabine Döring and Fritz Feger	
Emotions and Judgments About Risk	107
Robert C. Roberts	
The Moral Risks of Risky Technologies	127
Peter Goldie	
Ethical Imagination: Broadening Laboratory Deliberations	139
Simone van der Burg	
Part III Emotions as a Guide to Acceptable Risk	
Emotion in Risk Regulation: Competing Theories	159
Dan M. Kahan	

Emotions Within the Bounds of Pure Reason: Emotionality and Rationality in the Acceptance of Technological Risks 177
Dieter Birnbacher

Emotions Involved in Risk Perception: From Sociological and Psychological Risk Studies Towards a Neosentimentalist Meta-Ethics . . 195
Felicitas Kraemer

Risk Emotions and Risk Judgments: Passive Bodily Experience and Active Moral Reasoning in Judgmental Constellations 213
Mark Coeckelbergh

Emotional Reflection About Risks 231
Sabine Roeser

Index 245

Contributors

Dieter Birnbacher Department of Philosophy, Düsseldorf University, Düsseldorf, Germany, Dieter.Birnbacher@uni-duesseldorf.de

Ross Buck Communication Sciences, University of Connecticut, Storrs, CT, USA, ross.buck@uconn.edu

Mark Coeckelbergh Philosophy Department, Twente University, Enschede, The Netherlands, m.coeckelbergh@utwente.nl

Whitney A. Davis Davis Law Firm, Sacramento, CA, USA, wdavislaw@comcast.net

Ronald de Sousa Department of Philosophy, University of Toronto, Toronto, ON, Canada, Sousa@chass.utoronto.ca

Sabine Döring Philosophisches Seminar, Universität Tübingen, Tübingen, Germany, mail@sabinedoering.de

Dylan Evans Behavioural Science, School of Medicine, University College Cork, Cork, Ireland, evansd66@googlemail.com

Fritz Feger Philosophisches Seminar, Universität Tübingen, Tübingen, Germany, mail@fritzfeger.de

Peter Goldie Department of Philosophy, University of Manchester, Manchester, UK, peter.goldie@manchester.ac.uk

Dan M. Kahan Elizabeth K. Dollard Professor of Law, Yale Law School, New Haven, CT, USA, Dan.Kahan@yale.edu

Felicitas Kraemer Faculty of Innovation Management and Industrial Design, Philosophy & Ethics, Eindhoven University of Technology, Eindhoven, The Netherlands, f.Kraemer@tue.nl

Robert C. Roberts Department of Philosophy, Baylor University, Waco, TX, USA, Robert_Roberts@Baylor.edu

Sabine Roeser Department of Philosophy, Delft University of Technology, Delft, The Netherlands, s.roeser@tudelft.nl

Paul Slovic Decision Research and Department of Psychology, University of Oregon, Eugene, OR, USA, pslovic@darkwinguoregon.edu

Robert C. Solomon[†] was Quincy Lee Centennial Professor of Business and Philosophy and Distinguished Teaching Professor at the University of Texas at Austin, kmhiggins@mail.utexas.edu

Cass R. Sunstein Law School and Department of Political Science, University of Chicago, Chicago, IL, USA, csunstei@uchicago.edu

Simone van der Burg Philosophy Department, Twente University, Enschede, The Netherlands, S.vandenburg@utwente.nl

Emotions and Risky Technologies: Introduction and Overview

Sabine Roeser

Introduction

The risks involved in technologies such as cloning, GM-foods, and nuclear power plants spark emotional and heated debates. Many people are afraid of the unwanted possible consequences of such technologies. This gives rise to the following normative question: what role should emotions play when we judge whether a technology and its concomitant risks are morally acceptable? This question has direct practical implications: should engineers, scientists and policy makers involved in developing risk regulation take emotional responses (of the public, but also their own) seriously or not?

Though there is a great deal of empirical research on emotions and risky technologies, until now there has been almost no philosophical research on this topic. Some empirical psychologists think that emotional responses to risks are heuristics, creating biases that need to be corrected by rational and analytic procedures. However, many researchers studying emotion think that emotions are an essential part of practical rationality. So could emotions function as a normative guide when making judgments about morally acceptable risks?

This book covers new territory in that it sets the stage for research into the role of emotions in making moral decisions about risky technologies. It brings together leading scholars working in the fields of risk perception, emotions, and the ethics of risk and lets them reflect on this exciting and important new topic. Currently, the role of emotions in risk perception is studied by psychologists, sociologists and legal scholars. These scholars use an empirical approach, whereas most contributions in this book are philosophical (although they still engage with empirical work). In addition, many existing studies assume there is a clear distinction between reason and emotion, whereas most of the contributions in this book consider emotions as a source of practical rationality.

S. Roeser (✉)

Department of Philosophy, Delft University of Technology, Delft, The Netherlands
e-mail: s.roeser@tudelft.nl

I will first present a sketch of the theoretical background for the study of risk and emotion. I will then give an overview of the various contributions included in this book.

Background

Technological advances are inevitably accompanied by risks which raise important ethical issues that need to be dealt with by the societies that produce these technologies. Recent technologies such as nanotechnology, biotechnology, ICT, and nuclear power can improve human well-being but can also jeopardize our well-being through accidents or pollution, for example. Once such dangers become commonly known, technologies can trigger emotions, including fear and indignation, which often leads to conflict between experts and laypeople. How should we deal with such emotions in decision making about risky technologies?

Empirical research by Paul Slovic and his colleagues shows that emotions are a major determinant in risk perception (they term this phenomenon the “affect heuristic” or “risk as feeling”; Alhakami and Slovic 1994; Slovic 1999; Finucane et al. 2000; Slovic et al. 2002; recently, several journals have devoted special issues to this topic: *Risk Management* 2008, no. 3, *The Journal of Risk Research* 2006, no. 2). Many researchers who write on this topic assume that reason and emotion are distinct faculties (Dual Process Theory, cf. Epstein 1994; Sloman 1996, 2002; Stanovich and West 2002, cf. for similar views Haidt 2001; Greene and Haidt 2002; Prinz 2004; Greene 2007). Slovic writes that emotion and reason can interact and that we should take the emotions of the public seriously since they convey meaning (Slovic et al. 2004). Yet other scholars think that emotions are unreflective gut reactions that should be excluded from decision making about risk (Sunstein 2005); others think that emotions should at most be accepted as an unfortunate fact of life (Loewenstein et al. 2001, p. 281; Wolff 2006), while some think they can be used instrumentally, to increase the level of acceptance of a technology (De Hollander and Hanemaaijer 2003; Costa-Fond et al. 2008 illustrate this with the example of the acceptance of GM-food). The contributors to Part I of this book use Dual Process Theory to argue that emotions can bias our judgment of risk through an incorrect understanding of quantitative information.

However, many scholars argue that risk is not just a quantitative notion but that it also raises ethical considerations that are insufficiently addressed by conventional methods of risk assessment (Fischhoff et al. 1981; Shrader-Frechette 1991a; Krinsky and Golding 1992; Slovic 2000; Jaeger et al. 2001). The predominant approaches to risk assessment are based on cost-benefit analyses. Trade-offs between risks and benefits are unavoidable when we are judging the moral acceptability of risky technologies. There are no risk-free options; rejecting a technology also entails a risk (Sunstein 2005). For example, there is a slight risk that our house may collapse on us but without a house we would be vulnerable to the daily risks of the elements. Ideally, we try to maximize the benefits and minimize the risks of

a technology, but we can never rule out risk completely. Yet, it is not clear how we should balance the risks and benefits of technologies. Current approaches in risk assessment are generally based on technocratic methodologies that omit explicit reflection on ethical values. Various scholars have pointed out that technocratic approaches are far from value-neutral as these approaches make assumptions about which kinds of consequences matter and focus only on statistical information (cf. e.g. Fischhoff et al. 1981; Shrader-Frechette 1991a; Slovic 2000; Hansson 2009). Within the literature on acceptable risk, there is a consensus about which ethical considerations are relevant (empirically; e.g. Slovic 2000) and are important (normatively; e.g. Shrader-Frechette 1991a, contributions to Asveld and Roeser 2009). Examples are ethical considerations such as justice, fairness and autonomy.

In addition, it is doubtful whether there is a strict dichotomy between reason and emotion (Roeser 2009, 2010). Various philosophers and psychologists have argued that to be practically rational we need emotions (philosophers: e.g. de Sousa 1987; Greenspan 1988; Solomon 1993; Blum 1994; Little 1995; Stocker 1996; Goldie 2000; Ben Ze'ev 2000, psychologists: e.g. Scherer 1984; Frijda 1987; Lazarus 1991; Damasio 1994). According to some cognitive theories of emotions, emotions are judgments of value (Solomon 1993; Nussbaum 2001; Zagzebski 2003). Such theories help to see that emotions are necessary for ethical knowledge about risk (Shrader-Frechette 1991b; Bandes 2008; Roeser 2006, 2010, cf. the contributions to Parts II and III of this book).

To have well-grounded insight about whether a technological risk is morally acceptable, we need ethical intuitions and emotions (Roeser 2006, 2007). We need emotions such as sympathy to grasp ethical considerations such as justice, fairness and autonomy when making decisions about acceptable risk. For example, enthusiasm for a technology can signify that there are benefits for our well-being, whereas fear and worry can indicate that a technology is a threat to our well-being. Sympathy and empathy can give us a better insight into the just distribution of risks and benefits while indignation can reflect how technological risks that are imposed on us against our will violate our autonomy (Roeser 2006). This approach provides for a conceptual and normative framework that supports Slovic's claims about the important role played by the general public's emotions and intuitions (Roeser 2006, 2007). The emotions and intuitions of laypeople reflect a broader perception of risk that also includes important ethical considerations (Slovic 2000; Roeser 2007). Sunstein (2005) argues that emotions can lead to errors in risk perception such as "probability neglect". However, technocratic approaches to risk can lead to the omission of important ethical issues which results in "complexity neglect" (cf. Roeser, this book; also cf. Kahan and Paul 2006; Kahan et al. 2006).

Laypeople are often accused of being emotional and irrational in their judgments about, for example, nuclear energy, for which they are assured that the probability of a meltdown is statistically extremely low. However, laypeople often take Chernobyl as a point of reference. That accident was also statistically highly unlikely, and yet it happened, with catastrophic consequences. Proponents of nuclear energy should focus their arguments more on why a meltdown in a modern reactor would not have the same devastating consequences as the Chernobyl meltdown (Roeser 2006).

In addition, it can be quite rational to mistrust claims about low probability since scientists also err (cf. Shrader-Frechette 1991b; Hansson 2004). The emotions of laypeople can be based on reasonable concerns and it is therefore valid to address these in political debates on risk. In addition, emotions can play a role in risk perception of not only laypeople but also experts and politicians. These emotions should be taken seriously since they might reveal important ethical insights.

The aspect of emotion is usually ignored in risk management and if it is addressed it is applied instrumentally to create support for a technology. The contributions in Parts II and III of this book offer reasons for taking emotions seriously as part of moral decision making about risky technologies. Emotions should not be seen as the cul-de-sac of debates, but rather as the starting point for thorough ethical reflection. Of course people's emotional responses differ, but disagreement is nearly always a part of collective decision making, whether emotions are included or not. We should accept people's diverging emotional responses and discuss the concerns that underlie them. Considering diverging emotions and views will lead to more balanced judgments. Like other sources of knowledge, emotions can be misguided. Emotions have to be critically assessed, but in such a critical assessment, emotions should also play a role. Emotions are an important source of second order reflection about our first order emotions (Lacewing 2005). What are the concerns that inform the emotions of the general public, the experts and the policy makers? Those concerns need to be examined carefully and genuinely addressed in public debates. Political decision making procedures and modes of risk communication that take emotion seriously can address emotional responses by showing that certain considerations are unfounded. On the other hand, if it is found that emotional responses are indeed based on valid concerns, this should lead to a shift in risk management.

Overview of the Contributions

The three parts of this book contain contributions that address specific areas. Part I discusses the way emotions can distort risk perception. Most authors who write about risk and emotion invoke Dual Process Theory, according to which our thinking works in two different ways: system 1 is intuitive and emotional; system 2 is analytical. According to the scholars who adopt this approach, system 1 is fast but unreliable, whereas system 2 is normatively superior but slower than system 1. Emotions are seen as heuristics that can be useful but that are notoriously biased and then need to be corrected by analytical or quantitative methods. The contributing authors for Part I of this book conclude that in the light of these findings, a "liberal paternalist" (Sunstein) system forcing us to act appropriately is justified, rather than letting people rely on their highly unreliable emotions.

Chapter 1. In his contribution, *Cass Sunstein* builds on the seminal work of Tversky and Kahneman on heuristics and biases in probabilistic thinking, but he examines the domain of moral heuristics that has until now rarely been studied. Sunstein identifies a set of heuristics that influence widely-held factual and moral

judgments in the domain of risk. These heuristics work well until they are taken out of their context and are applied to domains for which they are not suitable. The first heuristic Sunstein discusses is “moral framing”. Just as in general framing, the way information is presented very much determines people’s judgments. Another heuristic is to condemn morally those who knowingly engage in acts that will result in human death. This is generally a useful heuristic, but according to Sunstein it fails in cases where death is the unintended side-effect of an overall desirable action. Another heuristic is that “People should not be permitted to engage in moral wrongdoing for a fee”. This heuristic is appropriate when correctly applied, but people apply it also to cases that according to Sunstein are not overtly morally wrong, such as activities that may produce pollutants as a byproduct. Yet another heuristic that works well in some cases but fails in others concerns so-called “betrayal risk”: it turns out that people are especially averse to risks that come from products designed to promote safety, such as airbags and vaccinations.

Chapter 2. *Ronald de Sousa* discusses various studies that indicate that our intuitive mode of thinking can mislead our assessment of risk and needs to be corrected by analytical methods. Emotions play a complex role in this context. According to de Sousa, emotions can be seen as mainly intuitive responses but they also play a role in more analytical responses. However, this does not mean that they inherit the advantages of both the systems described above (fast (intuitive) as well as reliable (analytical)), rather, it makes their status more dubious. According to de Sousa, we need emotions in determining our values. However, emotions are not always legitimate, such as a fear of terrorism which tends to be out of proportion and to feed upon itself. Resistance to nuclear technology, GM-food and nanotechnology is based on biased emotional responses that reinforce each other. If we compare emotions to one of the paradigm methods in rational decision theory, i.e. Bayesianism, there are four possible emotional biases to Bayesian decision making: emotions can bias our assessment of probabilities, of the values we attach to possible outcomes, of the product of both parameters or by bypassing a Bayesian calculation altogether. de Sousa distinguishes three stages of policy-making: (A) discovery, (B) justification and (C) motivation. According to de Sousa, whereas emotions are necessary in (A) and (C), they should not play a role in (B). Emotions need to be placed in a reflective equilibrium. Philosophers should contribute to a critical reflection on our emotions.

Chapter 3. In his contribution, *Paul Slovic* discusses how it is possible that people have such difficulties with responding appropriately to cases of mass tragedies, be it from genocide, natural disasters or as effects of use of technologies, such as global warming or nuclear weapons. An appropriate response would be to come to action, e.g. by intervening in the case of a genocide, by helping in the case of natural disasters or by changing our behavior in the case of global warming. However, we generally do not respond in such ways. Based on psychological studies, Slovic shows that statistics and numbers do not convey emotion but leave us “numbed by numbers”. Even though individual cases elicit feelings of sympathy and compassion that can induce us to take action, as numbers increase we respond less and less emotionally. One of the unsettling findings that Slovic presents is that this effect starts as soon as the number of victims is larger than one. One would think that the

more victims, the more compassion we would feel, but in fact, the opposite occurs. One personal story may affect us deeply, but as soon as more victims are involved, we become indifferent. The larger the numbers, the less impact an event has on us, culminating in an “utter collapse of compassion represented by apathy toward genocide” and other mass tragedies. Slovic concludes that we should not only rely on our ethical intuitions and emotions but that we should supply them with moral arguments and we should secure justice through legal and institutional mechanisms that compel us to act.

Chapter 4. *Ross Buck* and *Whitney Davis* show how the communication of risk fails because it does not address the emotions of the potential users of risky products. Warning labels are mainly directed at our analytic system, but to be effective, they should also be able to engage with our intuitive, emotional system. Marketing experts have long been aware of this. For example the tobacco and alcohol industries use emotionally based messages to entice people to engage in risky behaviors such as smoking and drinking by inducing a “mindless acceptance of risk”. The authors argue that risk communication experts should also make use of emotional response to elicit desirable behavior. They discuss various examples of warning labels that have proved largely ineffective because the role of emotions in risk perception has been underestimated. Buck and Davis refer to studies which claim that the more emotional the warning on cigarette packets, the more effective they are. Not only can the emotion of fear lead to avoiding cigarettes, but a moral emotion such as resentment can result in resistance to and followed by action against the tobacco industry.

Chapter 5. In his contribution, *Dylan Evans* criticizes approaches that see emotions as a source of ethical knowledge in assessing risky technologies. He argues that emotions are too unreliable and too easily manipulated to be assigned any normative weight in moral deliberation about risky technologies. Nevertheless, Evans thinks that understanding emotional responses can be enlightening in understanding debates about technological risk. Starting from a Humean theory that sees emotions as a source of subjective values, Evans argues that emotions can help us understand the values people assign to technological risks. However, once this process is understood risk assessments should not involve emotion but be based on statistics and cost-benefit analysis alone.

While the contributors to Part I of this book favor rational, analytical approaches and are cautious about taking emotions into account in risk assessment, the authors contributing to Parts II and III emphasise the importance of emotions in making value judgments and are critical of analysis that does not take account of emotions. The contributions in Part II are all in the tradition of virtue ethics: virtuous risk assessment requires emotions.

Chapter 6. *Sabine Döring* and *Fritz Feger* are critical of analytical approaches to risk assessment, specifically of rational decision theory. They argue that the so-called St. Petersburg Paradox is only paradoxical within the framework of rational choice-theory. Intuitively, there is no paradox, and Döring and Feger argue that our intuitions can be justified from a virtue-ethical account. They claim that a virtuous person who assesses risk has the appropriate emotions and can therefore perceive

non-inferentially the right thing to do. A virtuous risk assessor is neither a gambler nor a coward but makes the right judgment, not based on a theory or an algorithm but on direct insight. Döring and Feger then introduce what they call the “inverted St. Petersburg Paradox”. This paradox parallels the structure of a catastrophic risk such as a nuclear meltdown: a low probability but an (almost) infinitely bad outcome. The authors argue that in the inverted St. Petersburg Paradox, rational decision theory also fails and that we need a virtue-ethical approach in order to make the right decision. They conclude by emphasizing that concerning other, simpler decisions, rational decision theory has its merits, but it cannot do without a virtue-ethical approach.

Chapter 7. *Robert C. Roberts* views emotion as a kind of perception that, like sensory perception, can yield epistemic benefits *or* lead us astray. Roberts distinguishes four criteria for evaluating judgments: correctness, justification, experiential immediacy, and understanding. He shows how emotions have the potential to yield such benefits, but also to undermine them. He distinguishes two kinds of virtues: virtues that dispose us to have appropriate emotions, and virtues that “enable us to manage or transcend our emotions in the interest of correct judgments”. Roberts sees emotions as “concern-based construals”. “Concern” refers to the affective and motivational aspect of an emotion, “construal” to the “cognitive” aspect, though his view is an indirect criticism of any strict division between the cognitive and the non-cognitive, as Dual Process Theory seems to presuppose. He speaks of construal rather than belief because we do not always believe our emotions. An example is the person with a phobia of flying, who knows that flying is safer than other modes of transportation that the phobic doesn’t fear. Fear is the dominant emotional response to risky technologies. Fear can be a warranted and accurate perception, or it can be inaccurate. Roberts discusses the various ways fear of certain technologies can be misplaced. However, unlike the contributions in Part I, Roberts does not think that analytical methods are sufficient for forming risk judgments. He argues that “getting one’s risk-judgments right depends significantly on a correct emotional formation”.

Chapter 8. *Peter Goldie* discusses the uneasiness and ambivalence with which we relate to technological products such as computers, robots, avatars and other kinds of emotion-oriented technologies. Despite increased technological sophistication, we are pretty confident that such products do not have “phenomenal consciousness”, i.e. there is nothing it is like to be a computer, a robot or an avatar. Nevertheless, we can have negative emotional responses to such products treating them as if they were intentional agents, for example by getting angry and abusive towards them when they malfunction. The question arises whether we have moral obligations not to treat technological products in these ways. Goldie sees the main reason for a moral obligation not to abuse technological products in that this way of behaving can become habitual and transform into abusive behavior towards human beings. Drawing an analogy with Kant’s discussion of our duties with regard to non-human animals, Goldie argues that we have a duty to avoid this slippery slope, and to control our habits in such a way that does not prevent the proper treatment of other human beings.

Chapter 9. *Simone van der Burg* describes a case study that she has conducted as an “embedded ethicist”. As an ethicist, she was involved with a group of technical researchers who were developing a new technology. Van der Burg emphasizes the importance of including ethical reflection in the design phase of a technology, because there is then still a real chance that a technology can take ethical concerns into account. The technology she was involved with was an acousto-optic monitoring device to allow diabetes patients to detect their blood sugar-level in a non-invasive way. After studying the relevant background-literature, van der Burg noted that this device would probably only work properly on skin with little pigment. This could mean that the device would not be reliable for people with dark skin. Van der Burg pointed this out to the engineers but they were not as concerned about this problem as she was. Van der Burg believes that this reflected an inability to put themselves in the position of the potential users of the technology and how it could affect their quality of life. Diabetes patients with dark(er) skin also need new technologies with which they can monitor their blood sugar-levels without pain. Emotional capacities such as empathy and sympathy are needed to detect moral values that are entrenched in technologies.

The contributors in Part II argue that virtuous risk perception needs emotions. The contributors to Part III argue more generally that emotions are needed as a guide to judging the moral acceptability of risk.

Chapter 10. *Dan Kahan* explicitly rejects a so-called “irrational weigher”-theory. This is an often defended idea, which is also articulated by several contributors to Part I of this book. The claim is that emotions distort sound judgments about technological risks and should be exempt from decision making about risky technologies. In contrast, Kahan argues for what he calls the “cultural evaluator”-theory. This alternative theory is based on a different interpretation of the empirical evidence on the importance of emotions in risk perception. Supposed biases of emotions turn out to be reasonable attitudes if one presupposes a different theoretical framework. Kahan describes a study that he conducted with Paul Slovic and other scholars that shows that the more information people have about nanotechnology, the more affectively loaded are their assessments of its risks. Emotions are not separated from rational information as in Dual Process Theory. Rather than seeing emotions as heuristics that can shortcut supposedly more sophisticated forms of rational deliberation, Kahan argues that emotion functions as “a perceptive faculty uniquely suited to discerning what stance toward risk best coheres with a person’s values”. This claim has important normative implications: rather than leaving risk assessment to supposedly rational experts, the emotions and moral views of citizens should be included in decision making about risky technologies.

Chapter 11. *Dieter Birnbacher* discusses the stalemate between quantitative and qualitative approaches to risk assessment, the former proposing to leave risk assessment to scientific experts who use Bayesian decision theory, the latter proposing to include emotions and moral values expressed by lay people in their perception of risk. Birnbacher emphasizes that these approaches overlap to some degree. On the one hand, Bayesian approaches also include subjective preferences, emotional attitudes and value-judgments in their utility functions. On the other hand, defenders

of qualitative approaches are aware of the fact that laypeople's intuitions and emotions can be misguided. Birnbacher investigates how these two approaches can be brought even closer together, by including more emotional, qualitative considerations in Bayesian decision making than is presently the case. He argues that it is possible to include, for example, the following qualitative considerations into Bayesianism: quality of life, fairness, equality, the emotional damage caused by risk and the benefits of security, irreversibility, control, voluntariness and potential for catastrophe. However, Birnbacher thinks that other emotional-ethical considerations are not as easily integrated into Bayesianism. These are uncertainty about probabilities and/or possibly averse outcomes, and the fact that people have diverging risk perceptions that should be respected. Nevertheless, some emotional considerations should be excluded from risk assessment since they can be misleading. Examples are symbolic values, salience, familiarity, and naturalness. Birnbacher concludes that emotions are a mixed blessing: they can be sources of moral insight, but they can also mislead us.

Chapter 12. *Felicitas Kraemer* proposes a "neosentimentalist" account of risk perception. She starts her contribution with a short overview of the ideas of "objectivists" about risk, who suggest that risk is a purely quantitative, objective notion, and how this position has been proven to be untenable. Instead, the predominant view in the sociological and psychological literature on risk is a form of social constructivism: our understanding of risk depends to a large degree on our cultural background, contingent value judgments and emotions. Kraemer goes on to discuss Lennart Sjöberg's recent criticism of constructivist accounts and the role they assign to emotions. Sjöberg thinks that value judgments about risks are objective and rational and do not involve emotions. Kraemer questions Sjöberg's rationalist theory and thinks that his view of emotions as irrational gut reactions is too limited. Instead, she proposes an understanding of risk-emotions in line with neosentimentalist accounts found in metaethics such as that proposed by David Wiggins. According to a neosentimentalist account, values are constructions and projections of emotions, but they are not irrational and arbitrary. They can be assessed by criteria of rationality that hold within our culture. Kraemer thinks that such an account implies that emotions should play an important role in debates about risky technologies.

Chapter 13. Like the other contributors to Parts II and III of this book, *Mark Coeckelbergh* rejects the idea that the supposedly irrational emotions of ordinary people should be disregarded. However, Coeckelbergh proposes to understand emotions not as a form of cognition or judgment, but as possibly intimately related to judgments. He refers to this with the notion of "judgmental constellations" which is more able to encompass the passive, raw, physiological aspects of emotion than is possible in cognitive theories of emotion. According to Coeckelbergh, cognitive theories of emotions overemphasize the rational, cognitive aspects of emotions and ignore their involuntary, physiological phenomenology. The judgmental constellations account that Coeckelbergh proposes instead allows for a close interconnection between judgments and feelings without reducing the one to the other. According to Coeckelbergh, emotions are not necessary for moral judgments about risks, as several contributors to Parts II and III argue, but they do enhance the quality of

such judgments. He rejects the claim that laypeople's risk judgments diverge from those of experts because they are "biased", and instead proposes to examine the relationship between emotions and beliefs that laypeople have. Emotional responses to risky technologies teach us important lessons about the things we value. According to Coeckelbergh, the rejection of laypeople's emotions can itself be characterized as a bias.

Chapter 14. *Sabine Roeser* discusses the various biases that risk scholars attribute to emotions. Roeser argues that, on closer inspection, some of these biases are not actually biases while other biases are not really based on emotions. When it is clear that emotions bias our quantitative understanding of risk, they have to be corrected by scientific methods. It is important that scientists provide scientific information about risks in an emotionally accessible way to laypeople and politicians. However, there are also emotional biases that do not affect our quantitative understanding of risk, but affect our moral understanding of risk, such as the egoistic emotions involved in a NIMBY¹-response. Roeser argues that emotional biases to our moral understanding of risk have to be corrected by other emotions. Reflection on moral emotions should itself be based on emotions. Moral emotions that are specifically suitable for reflection are sympathy, empathy and compassion. These kinds of moral emotions help us to transcend our own, narrow point of view and provide us with a better understanding of moral values. Roeser proposes the following division of labor: scientists should provide us with quantitative information about risks, but moral deliberation about risks should explicitly involve emotions.

So where do the divergent views in all these contributions leave us? Are there basically two camps who cannot agree, the anti-emotion camp (Part I) and the pro-emotion camp (Parts II and III)? Or are the views defended in this book compatible? I think that despite differences in emphasis and theoretical outlook, the views presented are in many ways compatible. There is a consensus that emotions can distort the perception of the quantitative aspects of risk but emotions are necessary to grasp the moral aspects of risk. Some researchers, such as Evans, view the latter claim as being descriptive: emotions show us the values that people have. However, other authors see it as a normative claim: emotions are proper, even necessary epistemic tools with which to access the moral dimension of risky technologies. Emotions are not infallible, but neither are other kinds of perception. Several contributors to this book emphasize that emotions should not only be checked by reason, but also by other emotions. Moreover purely rational approaches will have to be supplemented by an emotion-based approach if we are to fully grasp the ethics of technological risk.²

¹"Not in my backyard".

²With thanks to Robert C. Roberts and Paul Slovic for their very helpful comments on a draft of this introduction, and to Petry Kievit at the NIAS (Netherlands Institute for Advanced Studies) for her extremely helpful editing work and comments.

References

- Alhakami, A. S., and P. Slovic. 1994. A psychological study of the inverse relationship between perceived risk and perceived benefit, *Risk Analysis* 14: 1085–1096.
- Asveld, L. and S. Roeser eds. 2009. *The Ethics of Technological Risk*. London: Earthscan.
- Bandes, S. A. 2008. Emotions, values, and the construction of risk. *Pennsylvania Law Review PENNumbra* 156: 421.
- Ben-Ze'ev, A. 2000. *The Subtlety of Emotions*. Cambridge, MA: MIT Press
- Blum, L. A. 1994. *Moral Perception and Particularity*. New York: Cambridge University Press.
- Costa-Font, J., E., Mossialos, and C., Rudisill. 2008. Are feelings of genetically modified food politically driven? *Risk Management* 8: 218–234.
- Damasio, A. 1994. *Descartes' Error*. New York: Putnam.
- de Sousa, R. 1987. *The Rationality of Emotions*. Cambridge, MA: MIT-Press.
- Epstein, S. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49(8): 709–724.
- Finucane, M., A., Alhakami, P., Slovic, and S. M., Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* 13: 1–17.
- Fischhoff, B., S., Lichtenstein, P., Slovic, S. L., Derby, and R., Keeney. 1981. *Acceptable Risk*. Cambridge: Cambridge University Press.
- Frijda, N. 1987. *The Emotions*. Cambridge: Cambridge University Press.
- Goldie, P. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.
- Greene, J. D. 2007. The secret joke of Kant's soul. In *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*. W. Sinnott-Armstrong ed., 2–79, Cambridge, MA: MIT Press.
- Greene, J. D., and J., Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6: 517–523.
- Greenspan, P. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification*. New York, London: Routledge.
- Haidt, J. 2001. The emotional dog and its rational tail. A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.
- Hansson, S. O. 2009. An agenda for the ethics of risk. In *The Ethics of Technological Risk*. L. Asveld, and S. Roeser, eds., 11–23, London: Earthscan.
- Hansson, S. O. 2004. Philosophical perspectives on risk. *Techné* 8: 10–35.
- De Hollander, A. E. M. and Hanemaaijer, A. H. eds. 2003. *Nuchter omgaan met risico's*. Bilthoven: RIVM.
- Jaeger, C. J., O., Renn, E. A., Rosa, and T., Weblar. 2001. *Risk, Uncertainty, and Rational Action*. London: Earthscan
- Kahan, D. M., P., Slovic, J., Gastil, and D., Braman. 2006. Fear of democracy: A cultural evaluation of Sunstein on risk. *Harvard Law Review* 119: 1071–1109.
- Kahan, D. M., and S., Paul. 2006. Cultural evaluations of risk: 'values' or 'blunders'? *Harvard Law Review* 119: 1110.
- Krimsky, S. and Golding, D. eds. 1992. *Social Theories of Risk*. Westport: Praeger Publishers.
- Lacewing, M. 2005. Emotional self-awareness and ethical deliberation. *Ratio* 18: 65–81.
- Lazarus, R. 1991. *Emotion and Adaptation*. New York: Oxford University Press.
- Little, M. O. 1995. Seeing and caring: The role of affect in feminist moral epistemology. *Hypatia* 10: 117–137.
- Loewenstein, G. F., E. U., Weber, C. K., Hsee, and N., Welch. 2001. Risk as feelings. *Psychological Bulletin* 127: 267–286.
- Nussbaum, M. 2001. *Upheavals of Thought*. Cambridge: Cambridge University Press.
- Prinz, J. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Roeser, S. 2010. Intuitions, emotions and gut feelings in decisions about risks: Towards a different interpretation of 'neuroethics'. *The Journal of Risk Research* 13: 175–190.

- Roeser, S. 2009. The relation between cognition and affect in moral judgments about risk. In *The Ethics of Technological Risks*. L. Asveld, and S. Roeser, eds., 182–201, London: Earthscan.
- Roeser, S. 2007. Ethical intuitions about risks. *Safety Science Monitor* 11: 1–30.
- Roeser, S. 2006. The role of emotions in judging the moral acceptability of risks. *Safety Science* 44: 689–700.
- Scherer, K. R. 1984. On the nature and function of emotion: A component process approach. In *Approaches to Emotion*. K. R. Scherer, and P. Ekman, eds., 293–317, Hillsdale, London: Lawrence Erlbaum Associates.
- Shrader-Frechette, K. 1991a. *Risk and Rationality*. Berkeley: University of California Press
- Shrader-Frechette, K. 1991b. Feelings, fear, and technological risk. In *The Presence of Feeling in Thought*. B. den Ouden, and M. Moen, eds., New York: Peter Lang.
- Slooman, S. A. 2002. Two systems of reasoning. In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich et al. eds., 379–396, Cambridge: Cambridge University Press.
- Slooman, S. A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin* 119: 3–22.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24: 311–322.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2002. The affect heuristic. In *Intuitive Judgment: Heuristics and Biases*. T. Gilovich, D. Griffin, and D. Kahneman, eds., 397–420, Cambridge: Cambridge University Press.
- Slovic, P. 2000. *The Perception of Risk*. London: Earthscan
- Slovic, P. 1999. Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis* 19: 689–701.
- Solomon, R. 1993. *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett.
- Stanovich, K. E., and R. F., West. 2002. Individual differences in reasoning: Implications for the rationality debate? In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich et al. eds., 421–440, New York: Cambridge University Press.
- Stocker, M. with Elizabeth Hegemann 1996. *Valuing Emotions*. Cambridge: Cambridge University Press.
- Sunstein, C. R. 2005. *Laws of Fear*. Cambridge: Cambridge University Press
- Wolff, J. 2006. Risk, fear, blame, shame and the regulation of public safety. *Economics and Philosophy* 22: 409–427.
- Zagzebski, L. 2003. Emotion and moral judgment. *Philosophy and Phenomenological Research* 66: 104–124.

Part I
Emotions as Distortions About Risk

Moral Heuristics and Risk

Cass R. Sunstein

1 Introduction

Pioneering the modern literature on heuristics in cognition, Amos Tversky and Daniel Kahneman contended that “people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations” (Tversky and Kahneman 1974, p. 1124). Intense controversy has developed over the virtues and vices of the heuristics, most of them “fast and frugal,” that play a role in many areas (see Gilovich et al. 2002; Gigerenzer and Todd 1999). But the relevant literature has only started to investigate the possibility that in the moral and political domain, people also rely on simple rules of thumb that often work well but that sometimes misfire (see Baron 1994, 1998; Messick 1993). In fact the central point seems obvious. Much of everyday morality consists of simple, highly intuitive rules that generally make sense but that fail in certain cases. It is wrong to lie or steal, but if a lie or a theft would save a human life, lying or stealing is probably obligatory. Not all promises should be kept. It is wrong to try to get out of a longstanding professional commitment at the last minute, but if your child is in the hospital, you may be morally required to do exactly that.

One of my major goals in this essay is to identify a set of heuristics that now influence factual and moral judgments in the domain of risk, and to try to make plausible the claim that some widely held practices and beliefs are a product of those heuristics. Often moral heuristics represent generalizations from a range of

C.R. Sunstein (✉)

Law School and Department of Political Science, University of Chicago, Chicago, IL, USA
e-mail: csunstei@uchicago.edu

In January 2009, Sunstein began work in the Obama Administration, later to be confirmed by the United States Senate as Administrator of the Office of Information and Regulatory Affairs. No work was done on this essay after Sunstein began government employment, and nothing said here represents an official position of the United States in any way.

problems for which they are indeed well-suited (see Baron 1994), and hence most of the time, such heuristics work well. The problem comes when the generalizations are wrenched out of context and treated as freestanding or universal principles, applicable to situations in which their justifications no longer operate. Because the generalizations are treated as freestanding or universal, their application seems obvious, and those who reject them appear morally obtuse, possibly even monstrous. I want to urge that the appearance is misleading and even productive of moral mistakes. There is nothing obtuse, or monstrous, about refusing to apply a generalization in contexts in which its rationale is absent.

Because Kahneman and Tversky were dealing with facts and elementary logic, they could demonstrate that the heuristics sometimes lead to errors. Unfortunately, that cannot easily be demonstrated here. In the moral and political domains, it is hard to come up with unambiguous cases where the error is both highly intuitive and on reflection uncontroversial – where people can ultimately be embarrassed about their own intuitions. Nonetheless, I hope to show that whatever one’s moral commitments, moral heuristics exist and indeed are omnipresent, adversely affecting our reactions to social risks.

2 Ordinary Heuristics and an Insistent Homunculus

2.1 Heuristics and Facts

The classic work on heuristics and biases deals not with moral questions but with issues of fact, often in the domain of risk and probability. In answering hard factual questions, those who lack accurate information use simple rules of thumb. How many words, in four pages of a novel, will have “ing” as the last three letters? How many words, in the same four pages, will have “n” as the second-to-last letter? Most people will give a higher number in response to the first question than in response to the second (Tversky and Kahneman 1984) – even though a moment’s reflection shows that this is a mistake. People err because they use an identifiable heuristic – the availability heuristic – to answer difficult risk-related questions. When people use this heuristic, they answer a question of probability by asking whether examples come readily to mind. How likely is a flood, an airplane crash, a traffic jam, a terrorist attack, or a disaster at a nuclear power plant? Lacking statistical knowledge, people try to think of illustrations. For people without statistical knowledge, it is far from irrational to use the availability heuristic; the problem is that this heuristic can lead to serious errors of fact, in the form of excessive fear of small risks and neglect of large ones.

Or consider the representativeness heuristic, in accordance with which judgments of probability are influenced by assessments of resemblance (the extent to which A “looks like” B). The representativeness heuristic is famously exemplified by people’s answers to questions about the likely career of a hypothetical woman named Linda, described as follows: “Linda is 31 years old, single, outspoken, and

very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations” (see Kahneman and Frederick 2002; Mellers et al. 2001). People were asked to rank, in order of probability, eight possible futures for Linda. Six of these were fillers (such as psychiatric social worker, elementary school teacher); the two crucial ones were “bank teller” and “bank teller and active in the feminist movement.”

More people said that Linda was less likely to be a bank teller than to be a bank teller and active in the feminist movement. This is an obvious mistake, a conjunction error, in which characteristics A and B are thought to be more likely than characteristic A alone. The error stems from the representativeness heuristic: Linda’s description seems to match “bank teller and active in the feminist movement” far better than “bank teller.” In an illuminating reflection on the example, Stephen Jay Gould observes that “I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me – ‘but she can’t just be a bank teller; read the description’” (Gould 1991, p. 469). Because Gould’s homunculus is especially inclined to squawk in the moral domain, I shall return to him.

2.2 Attribute Substitution and Prototypical Cases

What is a heuristic? Kahneman and Shane Frederick have suggested that heuristics are mental shortcuts used when people are interested in assessing a “target attribute” and when they substitute a “heuristic attribute” of the object, which is easier to handle (Kahneman and Frederick 2002). Heuristics therefore operate through a process of *attribute substitution*. The use of heuristics gives rise to intuitions about what is true (see Myers 2002), and these intuitions sometimes are biased, in the sense that they produce errors in a predictable direction. Consider the question whether more people die from suicides or homicides. Lacking statistical information, people might respond by asking whether it is easier to recall cases in either class (the availability heuristic). The approach is hardly senseless, but it might also lead to errors, a result of “availability bias” in the domain of risk perception (see Kuran and Sunstein 1999). Sometimes heuristics are linked to affect, and indeed affect has even been seen as a heuristic (Slovic et al. 2002); but attribute substitution is often used for factual questions that lack an affective component.

Similar mechanisms are at work in the domain of morality and risk. Unsure what to think or do about a target attribute (what morality requires, what society should do about risks), people might substitute a heuristic attribute instead – asking, for example, about the view of trusted authorities (a leader of the preferred political party, an especially wise judge, a religious figure). Often the process works by appeal to *prototypical cases*. Confronted by a novel and difficult problem, observers often ask whether it shares features with a familiar problem. If it seems to do so, then the solution to the familiar problem is applied to the novel and difficult one. Of course it is

possible that in the domain of values as well as facts, real-world heuristics generally perform well in the real world - so that moral errors are reduced, not increased, by their use, at least compared to the most likely alternatives (see my remarks on rule-utilitarianism below). The only claim here is that some of the time, our moral judgments can be shown to misfire.

The principal heuristics should be seen in light of dual-process theories of cognition (Kahneman and Frederick 2002). Those theories distinguish between two families of cognitive operations, sometimes labeled System I and System II. System I is intuitive; it is rapid, automatic, and effortless (and it features Gould's homunculus). System II, by contrast, is reflective; it is slower, self-aware, calculative, and deductive. System I proposes quick answers to problems of judgment and System II operates as a monitor, confirming or overriding those judgments. Consider, for example, someone who is flying from New York to London in the month after an airplane crash. This person might make a rapid, barely conscious judgment, rooted in System I, that the flight is quite risky; but there might well be a System II override, bringing a more realistic assessment to bear. System I often has an affective component, but it need not; for example, a probability judgment might be made quite rapidly and without much affect at all.

There is growing evidence that people often make automatic, largely unreflective moral judgments, for which they are sometimes unable to give good reasons (see Greene and Haidt 2002; Haidt 2001; compare Pizarro and Bloom 2003). Moral, political, or legal judgments often substitute a heuristic attribute for a target attribute; System I is operative here as well, and it may or may not be subject to System II override. Consider the incest taboo. People have moral revulsion against incest even in circumstances in which the grounds for that taboo seem to be absent; they are subject to "moral dumbfounding" (Haidt et al. 2004), that is, an inability to give an account for a firmly held intuition. It is plausible, at least, to think that System I is driving their judgments, without System II correction. The same is true in legal and political contexts as well.

3 Heuristics and Morality

To show that heuristics operate in the moral domain, we have to specify some benchmark by which we can measure moral truth. On these questions I want to avoid any especially controversial claims. Whatever's one view of the foundations of moral and political judgments, I suggest, moral heuristics are likely to be at work in practice.

Many utilitarians, including John Stuart Mill and Henry Sidgwick, argue that ordinary morality is based on simple rules of thumb that generally promote utility but that sometimes misfire (see Mill 1971, pp. 28–29; Sidgwick 1981, pp. 199–216; originally published 1907; Hare 1981; Smart 1973). For example, Mill emphasizes that human beings "have been learning by experience the tendencies of experience," so that the "corollaries from the principle of utility" are being progressively captured

by ordinary morality (Mill 1971, p. 29).¹ Is ordinary morality a series of heuristics for what really matters, which is utility?

These large debates are not easy to resolve, simply because utilitarians and deontologists are most unlikely to be convinced by the suggestion that their defining commitments are mere heuristics. Here there is a large difference between moral heuristics and the heuristics uncovered in the relevant psychological work, where the facts or simple logic provide a good test whether people have erred. If people tend to think that more words, in a given space, end with the letters “ing” than have “n” in the next-to-last position, something has clearly gone wrong. If people think that some person Linda is more likely to be “a bank teller who is active in the feminist movement” than a “bank teller,” there is an evident problem. In the moral domain, factual blunders and simple logic do not provide such a simple test.

My goal here is therefore not to show, with Sigdwick and Mill, that common sense morality is a series of heuristics for the correct general theory, but more cautiously that in many particular cases, moral heuristics are at work – and that this point can be accepted by people with diverse general theories, or with grave uncertainty about which general theory is correct. In the cases catalogued below, I contend that it is possible to conclude that a moral heuristic is at work without accepting any especially controversial normative claims. In several of the examples, that claim can be accepted without accepting any contestable normative theory at all. Other examples will require acceptance of what I shall call “weak consequentialism,” in accordance with which the social consequences of the legal system are relevant, other things being equal, to what law ought to be doing.

Of course some deontologists will reject any form of consequentialism altogether. They might believe, for example, that retribution is the proper theory of punishment, and that the consequences of punishment are never relevant to the proper level of punishment. Some of my examples will be unpersuasive to deontologists who believe that consequences do not matter at all. But weak consequentialism seems to me sufficiently nonsectarian, and attractive to sufficiently diverse people, to make plausible the idea that in the cases at hand, moral heuristics are playing a significant role. And for those who reject weak consequentialism, it might nonetheless be productive to ask whether, from their own point of view, certain rules of morality and law are reflective of heuristics that sometimes produce serious errors.

4 The Asian Disease Problem and Moral Framing

In a finding closely related to their work on heuristics, Kahneman and Tversky themselves find “moral framing” in the context of what has become known as “the Asian

¹ On a widely held view, a primary task of ethics is to identify the proper general theory and to use it to correct intuitions in cases in which they go wrong (Hooker 2000). Consider here the provocative claim that much of everyday morality, nominally concerned with fairness, should be seen as a set of heuristics for the real issue, which is how to promote utility (see Baron 1998; to the same general effect, with numerous examples from law, see Kaplow and Shavell 2003).

disease problem” (Kahneman and Tversky 1984). Framing effects do not involve heuristics, but because they raise obvious questions about the rationality of moral intuitions, they provide a valuable backdrop. Here is the first component of the problem:

Imagine that the US is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Which of the two programs would you favor?

Most people choose Program A.

But now consider the second component of the problem, in which the same situation is given but followed by this description of the alternative programs:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Most people choose Problem D. But a moment’s reflection should be sufficient to show that Program A and Program C are identical, and so too for Program B and Program D. These are merely different descriptions of the same programs. The purely semantic shift in framing is sufficient to produce different outcomes. Apparently people’s moral judgments about appropriate programs depend on whether the results are described in terms of “lives saved” or instead “lives lost.” What accounts for the difference? The most sensible answer begins with the fact that human beings are pervasively averse to losses (hence the robust cognitive finding of loss aversion, Tversky and Kahneman 1991). With respect to either self-interested gambles or fundamental moral judgments, loss aversion plays a large role in people’s decisions. But what counts as a gain or a loss depends on the baseline from which measurements are made. Purely semantic reframing can alter the baseline and hence alter moral intuitions (for many examples involving fairness, see Kahneman et al. 1986).

Moral framing has been demonstrated in the important context of obligations to future generations (see Frederick 2003), a much-disputed question of morality, politics, and law (Revesz 1999; Morrison 1998). To say the least, the appropriate discount rate for those yet to be born is not a question that most people have pondered, and hence their judgments are highly susceptible to different frames. From a series of surveys, Maureen Cropper and her coauthors (1994) suggest that people are indifferent between saving one life today and saving 45 lives in 100 years. They make this suggestion on the basis of questions asking people whether they would choose a program that saves “100 lives now” or a program that saves a substantially larger number “100 years from now.” It is possible, however, that people’s responses depend on uncertainty about whether people in the future will otherwise die (perhaps technological improvements will save them?); and other ways of framing the same problem yield radically different results (Frederick 2003). For example, most

people consider “equally bad” a single death from pollution next year and a single death from pollution in 100 years. This finding implies no preference for members of the current generation. The simplest conclusion is that people’s moral judgments about obligations to future generations are very much a product of framing effects (for a similar result, see Baron 2000).²

The same point holds for the question whether government should consider not only the number of “lives” but also the number of “life years” saved by regulatory interventions. If the government focuses on life-years, a program that saves children will be worth far more attention than a similar program that saves senior citizens. Is this immoral? People’s intuitions depend on how the question is framed (see Sunstein 2004). People will predictably reject an approach that would count every old person as worth “significantly less” than what every young person is worth. But if people are asked whether they would favor a policy that saves 105 old people or 100 young people, many will favor the latter, in a way that suggests a willingness to pay considerable attention to the number of life-years at stake.

At least for unfamiliar questions of morality, politics, and law, people’s intuitions are very much affected by framing. Above all, it is effective to frame certain consequences as “losses” from a status quo; when so framed, moral concern becomes significantly elevated. It is for this reason that political actors often phrase one or another proposal as “turning back the clock” on some social advance. The problem is that for many social changes, the framing does not reflect social reality, but is simply a verbal manipulation.

Let us now turn to examples that are more controversial.

5 Morality and Risk Regulation

My principal interest here is the relationship between moral heuristics and questions of law and policy. The catalogue is meant to be illustrative rather than exhaustive.

5.1 *Cost-Benefit Analysis*

An automobile company is deciding whether to take certain safety precautions for its cars. In deciding whether to do so, it conducts a cost-benefit analysis, in which it

² Here too the frame may indicate something about the speaker’s intentions, and subjects may be sensitive to the degree of certainty in the scenario (assuming, for example, that future deaths may not actually occur). While strongly suspecting that these explanations are not complete (see Frederick 2003), I mean not to reject them, but only to suggest the susceptibility of intuitions to frames (for skeptical remarks, see Kamm 1998).

concludes that certain precautions are not justified – because, say, they would cost \$100 million and save only four lives, and because the company has a “ceiling” of \$10 million per lives saved (a ceiling that is, by the way, significantly higher than the amount the United States Environmental Protection Agency uses for a statistical life). How will ordinary people react to this decision? The answer is that they will not react favorably (see Viscusi 2000, pp. 547, 558). In fact they tend to punish companies that base their decisions on cost-benefit analysis, even if a high valuation is placed on human life. By contrast, they impose less severe punishment on companies that are willing to impose a “risk” on people but that do not produce a formal risk analysis that measures lives lost and dollars, and trades one against another (see Viscusi 2000; Tetlock 2000). The oddity here is that under tort law, it is unclear that a company should not be liable at all if it has acted on the basis of a competent cost-benefit analysis; such an analysis might even insulate a company from a claim of negligence. What underlies people’s moral judgments, which are replicated in actual jury decisions (Viscusi 2000)?

It is possible that when people disapprove of trading money for lives, they are generalizing from a set of moral principles that are generally sound, and even useful, but that work poorly in some cases. Consider the following moral principle: *Do not knowingly cause a human death*. In ordinary life, you should not engage in conduct with the knowledge that several people will die as a result. If you are playing a sport or working on your yard, you ought not to continue if you believe that your actions will kill others. Invoking that idea, people disapprove of companies that fail to improve safety when they are fully aware that deaths will result. By contrast, people do not disapprove of those who fail to improve safety while believing that there is a “risk” but appearing not to know, for certain, that deaths will ensue. When people object to risky action taken after cost-benefit analysis, it seems to be partly because that very analysis puts the number of expected deaths squarely “on screen” (see Tetlock 2000).

Companies that fail to do such analysis, but that are aware that a “risk” exists, do not make clear, to themselves or to anyone else, that they caused deaths with full knowledge that this was what they were going to do. People disapprove, above all, of companies that cause death knowingly. There may be a kind of “cold-heart heuristic” here: Those who know that they will cause a death, and do so anyway, are regarded as cold-hearted monsters.³ On this view, critics of cost-benefit analysis should be seen as appealing to System I and as speaking directly to the homunculus: “is a corporation or public agency that endangers us to be pardoned for its sins once it has spent \$6.1 million per statistical life on risk reduction?” (Ackerman and Heinzerling 2004).

Note that it is easy to reframe a probability as a certainty and vice-versa; if I am correct, the reframing is likely to have large effects. Consider two cases:

- (a) Company A knows that its product will kill ten people. It markets the product to its ten million customers with that knowledge. The cost of eliminating the risk would have been \$100 million.

³ I am grateful to Jonathan Haidt for this suggestion.

- (b) Company B knows that its product creates a 1 in 1 million risk of death. Its product is used by ten million people. The cost of eliminating the risk would have been \$100 million.

I have not collected data, but I am willing to predict that Company A would be punished more severely than Company B, even though there is no difference between the two.

I suggest, then, that a moral heuristic is at work, one that imposes moral condemnation on those who knowingly engage in acts that will result in human deaths.

And of course this heuristic does a great deal of good. The problem is that it is not always unacceptable to cause death knowingly, at least if the deaths are relatively few and an unintended byproduct of generally desirable activity. When government allows new highways to be built, it knows that people will die on those highways; when government allows new coal-fired power plants to be built, it knows that some people will die from the resulting pollution; when companies produce tobacco products, and when government does not ban those products, hundreds of thousands of people will die; the same is true for alcohol. Of course it would make sense, in all of these domains, to take extra steps to reduce risks. But that proposition does not support the implausible claim that we should disapprove, from the moral point of view, of any action taken when deaths are foreseeable.

There is a complementary possibility, involving the confusion between the ex ante and ex post perspective. If a life might have been saved by a \$50 expenditure on a car, people are going to be outraged, and they will impose punishment. What they will not see or incorporate is the fact, easily perceived ex ante, that the \$50-per-car expenditure would have been wasted on millions of other people. It is hardly clear that the ex ante perspective is always preferable. But something has gone badly wrong if the ex post perspective leads people to neglect the tradeoffs that are actually involved.

I believe that it is impossible to vindicate, in principle, the widespread social antipathy to cost-benefit balancing.⁴ But here too, “a little homunculus in my head continues to jump up and down, shouting at me” that corporate cost-benefit analysis, trading dollars for a known number of deaths, is morally unacceptable. The voice of the homunculus, I am suggesting, is not reflective, but instead a product of System I, and a crude but quite tenacious moral heuristic.

5.2 Emissions Trading

In the last decades, those involved in enacting and implementing environmental law have experimented with systems of “emissions trading” (Sunstein 2002). In those

⁴ I put to one side cases in which those who enjoy the benefits are wealthy and those who incur the costs are poor; in some situations, distributional considerations will justify a departure from what would otherwise be compelled by cost-benefit analysis (on this and other problems with cost-benefit analysis, see Sunstein 2002).

systems, polluters are typically given a license to pollute a certain amount, and the licenses can be traded on the market. The advantage of emissions trading systems is that if they work well, they will ensure emissions reductions at the lowest possible cost.

Is emissions trading immoral? Many people believe so. Political theorist Michael Sandel, for example, urges that trading systems “undermine the ethic we should be trying to foster on the environment” (Sandel 1997; see also Kelman 1981). Sandel contends:

[T]urning pollution into a commodity to be bought and sold removes the moral stigma that is properly associated with it. If a company or a country is fined for spewing excessive pollutants into the air, the community conveys its judgment that the polluter has done something wrong. A fee, on the other hand, makes pollution just another cost of doing business, like wages, benefits and rent.

In the same vein, Sandel objects to proposals to open carpool lanes to drivers without passengers who are willing to pay a fee. Here, as in the environmental context, it seems unacceptable to permit people to do something that is morally wrong so long as they are willing to pay for the privilege.

I suggest that like other critics of emissions trading programs, Sandel is using a moral heuristic; in fact he has been fooled by his homunculus. The heuristic is this: *People should not be permitted to engage in moral wrongdoing for a fee.* You are not allowed to assault someone so long as you are willing to pay for the right to do so; there are no tradable licenses for rape, theft, or battery. The reason is that the appropriate level of these forms of wrongdoing is zero (putting to one side the fact that enforcement resources are limited; if they were unlimited, we would want to eliminate, not merely to reduce, these forms of illegality). But pollution is an altogether different matter. At least some level of pollution is a byproduct of desirable social activities and products, including automobiles and power plants. Of course certain acts of pollution, including those that violate the law or are unconnected with desirable activities, are morally wrong; but the same cannot be said of pollution as such. When Sandel objects to emissions trading, he is treating pollution as equivalent to a crime in a way that overgeneralizes a moral intuition that makes sense in other contexts. There is no moral problem with emissions trading as such. The insistent objection to emissions trading systems stems from a moral heuristic.

Unfortunately, that objection has appeared compelling to many people, so much as to delay and to reduce the use of a pollution reduction tool that is, in many contexts, the best available (Sunstein 2002). Here, then, is a case in which a moral heuristic has led to political blunders, in the form of policies that impose high costs for no real gain.

5.3 Betrayals and Betrayal Risk

To say the least, people do not like to be betrayed. A betrayal of trust is likely to produce a great deal of outrage. If a babysitter neglects a child or if a security

guard steals from his employer, people will be angrier than if the identical acts are performed by someone in whom trust has not been reposed. So far, perhaps, so good: When trust is betrayed, the damage is worse than when an otherwise identical act has been committed by someone who was not a beneficiary of trust. And it should not be surprising that people will favor greater punishment for betrayals than for otherwise identical crimes (see Koehler and Gershoff 2003). Perhaps the disparity can be justified on the ground that the betrayal of trust is an independent harm, one that warrants greater deterrence and retribution – a point that draws strength from the fact that trust, once lost, is not easily regained. A family robbed by its babysitter might well be more seriously injured than a family robbed by a thief. The loss of money is compounded and possibly dwarfed by the violation of a trusting relationship. The consequence of the violation might also be more serious. Will the family ever feel entirely comfortable with babysitters? It is bad to have an unfaithful spouse, but it is even worse if the infidelity occurred with your best friend, because that kind of infidelity makes it harder to have trusting relationships with friends in the future.

In this light it is possible to understand why betrayals produce special moral opprobrium and (where the law has been violated) increased punishment. But consider a finding that is much harder to explain: *People are especially averse to risks of death that come from products (like airbags) designed to promote safety* (Koehler and Gershoff 2003). The aversion is so great that people have been found to prefer a higher chance of dying, as a result of accidents from a crash, to a significantly lower chance of dying in a crash as a result of a malfunctioning airbag. The relevant study involved two principal conditions. In the first, people were asked to choose between two equally priced cars, Car A and Car B. According to crash tests, there was a 2% chance that drivers of Car A, with Air Bag A, will die in serious accidents as a result of the impact of the crash. With Car B, and Air Bag B, there was a 1% chance of death, but also an additional chance of 1 in 10,000 (0.01%) of death as a result of deployment of the air bag. Similar studies involved vaccines and smoke alarms.

The result was that most participants (over two-thirds) chose the higher risk safety option when the less risky one carried a “betrayal risk.” A control condition demonstrated that people were not confused about the numbers: when asked to choose between a 2% risk and a 1.01% risk, people selected the 1.01% risk so long as betrayal was not involved. In other words, people’s aversion to betrayals is so great that they will increase their own risks rather than subject themselves to a (small) hazard that comes from a device that is supposed to increase safety. “Apparently, people are willing to incur greater risks of the very harm they seek protection from to avoid the mere possibility of betrayal” (Koehler and Gershoff 2003, p. 244). Remarkably, “betrayal risks appear to be so psychologically intolerable that people are willing to double their risk of death from automobile crashes, fires, and diseases to avoid incurring a small possibility of death by safety device betrayal.”

What explains this seemingly bizarre and self-destructive preference? I suggest that a heuristic is at work: *Punish, and do not reward, betrayals of trust*. The heuristic generally works well. But it misfires in some cases, as when those who deploy it

end up increasing the risks they themselves face. An airbag is not a security guard or a babysitter, endangering those whom they have been hired to protect. It is a product, to be chosen if and only if it decreases aggregate risks. If an airbag makes people safer on balance, it should be used, even if in a tiny percentage of cases it will create a risk that would not otherwise exist. Of course it is true that some kinds of death are reasonably seen as worse than others. It is not absurd to prefer one kind of death to another. But betrayal aversion is not adequately explained in these terms; the experimental work suggests that people are generalizing from a heuristic.

In a sense, the special antipathy to betrayal risks might be seen to involve not a moral heuristic but a taste. In choosing products, people are not making pure moral judgments; they are choosing what they like best, and it just turns out that a moral judgment, involving antipathy to betrayals, is part of what they like best. It would be useful to design a purer test of moral judgments, one that would ask people not about their own safety but about that of others – for example, whether people are averse to betrayal risks when they are purchasing safety devices for their friends or family members. There is every reason to expect that it would produce substantially identical results to those in the experiments just described. Closely related experiments support that expectation (see Ritov and Baron 2002, p. 168). In deciding whether to vaccinate their children from risks for serious diseases, people show a form of “omission bias.” Many people are more sensitive to the risk of the vaccination than to the risk from diseases – so much so that they will expose their children to a greater risk from “nature” than from the vaccine. (There is a clear connection between omission bias and trust in nature and antipathy to “playing God.”) as discussed below. The omission bias, I suggest, is closely related to people’s special antipathy to betrayals. It leads to moral errors, in the form of vaccination judgments, and undoubtedly others, by which some parents increase the fatality risks faced by their own children.

6 Conclusion

To the extent that moral heuristics operate as rules, they might be defended in the way that all rules are – better than the alternatives even if productive of error in imaginable cases. Moral heuristics might show a kind of “ecological rationality,” working well in most real-world contexts (Gigerenzer 2000); recall the possibility that human beings live by simple heuristics that make us good. My suggestion is not that the moral heuristics, in their most rigid forms, are socially worse than the reasonable alternatives. It is hard to resolve that question in the abstract. I am claiming only that such heuristics lead to real errors and significant confusion in thinking about risk. Regulators, after all, are not in the position of ordinary people, with limited time and in need of a simple rule of thumb. They typically have significant resources, including significant time, and they can do far better than to rely on heuristics.

If it is harder to demonstrate that heuristics are at work in the domain of morality than in the domain of facts, this is largely because we are able to agree, in the relevant cases, about what constitutes factual error, and often less able to agree about what constitutes moral error. With respect to the largest disputes about what morality requires, it may be too contentious to argue that one side is operating under a heuristic, whereas another side has it basically right. But I hope that I have said enough to show that in particular cases, sensible rules of thumb lead to demonstrable errors not merely in probability judgments, but in moral assessments of risks as well.

Acknowledgments I am grateful to Daniel Kahneman and Martha Nussbaum for valuable discussions. For helpful comments on a previous draft, I also thank participants in a seminar at Cambridge University, Jonathan Baron, Mary Anne Case, Elizabeth Emens, Robert Frank, Jonathan Haidt, Robert Goodin, Steven Pinker, Edward Stein, and Peter Singer. Some of the discussion here draws on Moral Heuristics, *Behavioral and Brain Sciences* 28: 531–546 (2005)

References

- Ackerman, F., and L., Heinzerling. 2004. *Priceless: On Knowing the Price of Everything and the Value of Nothing*. New York: The New Press.
- Baron, J. 1994. Nonconsequentialist decisions. *Behavioral and Brain Sciences* 17: 1–10.
- Baron, J. 1998. *Judgment Misguided: Intuition and Error in Public Decision Making*. Oxford: Oxford University Press.
- Baron, J. 2000. Can we use human judgments to determine the discount rate? *Risk Analysis* 20: 861–868.
- Cropper, M. L., S. K., Aydede, and P. R., Portney. 1994. Preferences for life-saving programs: How the public discounts time and age. *Journal of Risk and Uncertainty* 8: 243–265.
- Darley, J. M., K. M., Carlsmith, and P. H., Robinson. 2000. Incapacitation and just deserts as motives for punishment. *Law and Human Behavior* 24: 659–683.
- Frederick, S. 2003. Measuring intergenerational time preference: Are future lives valued less? *Journal of Risk and Uncertainty* 26: 39–53.
- Gigerenzer, G. 2000. *Adaptive Thinking: Rationality in the Real World*. Oxford: Oxford University Press.
- Gigerenzer, G., and P., Todd. 1999. *Simple Heuristics that Make us Smart*. Oxford: Oxford University Press.
- Gilovich, T., D. Griffin, and D. Kahneman, eds., 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.
- Gould, S. J. 1991. *Bully for Brontosaurus: Reflections in Natural History*. New York: W.W. Norton and Company.
- Greene, J., and J., Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6: 517–523.
- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.
- Haidt, J., F., Bjorklund, and S., Murphy. 2004. *Moral Dumbfounding: When Intuition Finds No Reason*. Unpublished manuscript: University of Virginia.
- Hare, R. M. 1981. *Moral Thinking*. Oxford: Oxford University Press.
- Hooker, B. 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.

- Kahneman, D., and S., Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, D. Griffin, and D. Kahneman, eds., Cambridge: Cambridge University Press.
- Kahneman, D., J. L., Knetsch, and R. H., Thaler. 1986. Fairness as a constraint on profit-seeking: Entitlements in the market. *American Economic Review* 76: 728–741.
- Kahneman, D., and A., Tversky. 1984. Choices, values, and frames. *American Psychologist* 39: 341–350.
- Kamm, F. 1998. Moral intuitions, cognitive psychology, and the harming-versus-not-aiding distinction. *Ethics* 108: 463–488.
- Kaplow, L., and S., Shavell. 2003. *Fairness Versus Welfare*. Cambridge, MA: Harvard University Press.
- Kelman, S. 1981. *What Price Incentives? Economists and the Environment*. Boston: Auburn House.
- Koehler, J. J., and A. D., Gershoff. 2003. Betrayal aversion: When agents of protection become agents of harm. *Organizational Behavior and Human Decision Processes* 90: 244–261.
- Kuran, T., and C. R., Sunstein. 1999. Availability cascades and risk regulation. *Stanford Law Review* 51: 683–768.
- Mellers, B., R., Hertwig, and D., Kahneman. 2001. Do frequency representations eliminate conjunction effects? *Psychological Science* 12: 269–275.
- Messick, D. 1993. Equality as a decision heuristic. In *Psychological Perspectives on Justice*. B. Mellers, and J. Baron, eds., Cambridge: Cambridge University Press.
- Mill, J. S. 1971. *Utilitarianism*. New York: Bobbs-Merrill Company.
- Morrison, E. R. 1998. Comment: Judicial review of discount rates used in regulatory cost-benefit analysis. *University of Chicago Law Review* 65: 1333–1370.
- Myers, D. G. 2002. *Intuition: Its Powers and Perils*. New Haven: Yale University Press.
- Pizarro, D. A., and P., Bloom. 2003. The intelligence of the moral intuitions: Comment on Haidt. *Psychological Review* 110: 193–198.
- Revesz, R. 1999. Environmental regulation, cost-benefit analysis, and the discounting of human lives. *Columbia Law Review* 99: 941–1017.
- Ritov, I., and J., Baron. 2002. Reluctance to vaccinate: Omission bias and ambiguity. In *Behavioral Law and Economics*. C. R. Sunstein, ed., Cambridge: Cambridge University Press.
- Sandel, M. 1997. *It's Immoral to Buy the Right to Pollute*, N.Y. TIMES, December 15, 1997: A23.
- Sigdwick, H. 1981, (originally published 1907). *The Methods of Ethics*. Indianapolis: Hackett Publishing Company.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2002. The affect heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, D. Griffin, and D. Kahneman, eds., Cambridge: Cambridge University Press.
- Smart, J. J. C. 1973. An outline of a system of utilitarian ethics. In *Utilitarianism: For and Against*. J. J. C. Smart, and B. Williams, eds., Cambridge: Cambridge University Press.
- Sunstein, C. R. 2002. *Risk and Reason. Safety, Law, and the Environment*. Cambridge: Cambridge University Press.
- Sunstein, C. R. 2004. Lives, life-years, and willingness to pay. *Columbia Law Review* 104: 205–252.
- Tetlock, P. 2000. Coping with tradeoffs. In *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. A. Lupia, S. Popkin, and M. D. McCubbins, eds., Cambridge: Cambridge University Press.
- Tversky, A., and D., Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.
- Tversky, A., and D., Kahneman. 1984. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review* 90: 293–315.
- Tversky, A., and D., Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 106: 1039–1061.
- Viscusi, W. K. 2000. Corporate risk analysis: A reckless act? *Stanford Law Review* 52: 547–597.

Here's How I Feel: Don't Trust Your Feelings!

Ronald de Sousa

1 The Ambiguity of "Risk"

The simplest understanding of the concept of risk is as "the probability of a dangerous event ($p(E)$) multiplied by the amount of the expected damage (D) connected to this event: $R(E) = p(E) \times D$ " (Bora 2007). In common speech and practice, however, that clear concept quickly becomes murky as talk of risk appears to reflect a confusing multiplicity of meanings.

For a start, it refers to at least two distinct aspects of a situation: the nature of bad consequences that might follow, or the likelihood of their occurrence. In "The main risk involved in rollerblading is injury from collision with cars," the former seems intended, while "There is a risk of death but it is low" suggests the latter. Furthermore, the perception and response to danger, including the affective response of fear, affords one of the clearest illustrations of the "two track" view of brain functioning. *Intuitive*, evolutionarily more ancient "First Track" processes are rapid, generally unconscious, and typically manifested in immediate emotional responses. The *Analytic* or "Second Track" processes are explicit, language-driven inferences that work in parallel but not always in harmony with the Intuitive system.¹ The main goal of the present essay is to sketch some consequences of these complexities in the concept of risk, and of the dual origins of our responses. My central thesis is that while we cannot avoid grounding our assessments in emotion, we should regard them with extreme skepticism. Objectivity, in the sense of inter-subjective and multimodal consilience, remains an ideal worth striving for in the perception of danger in general, and of risks posed by technology in particular.

R. de Sousa (✉)

Department of Philosophy, University of Toronto, Toronto, ON, Canada
e-mail: Sousa@chass.utoronto.ca

¹What I refer to as the "intuitive" track is more or less equivalent to what Paul Slovic calls the "perceptual" system (Slovic et al. 2004). There are many formulations of the basic distinction, notably Strack and Deutsch (2004), who use the terms "impulsive" and "reflective", and Stanovich (2004), who cites some two dozen other versions of the idea of the two-track mind.

2 Two Standard Models of Decision

The main point of assessing risk in practice is to guide the actions and decisions we take in response to it. It makes sense, then, to begin with some remarks about how we should understand the concepts of “action” and “decision”.

Any action or decision is undertaken in the light of beliefs about the agent’s current situation, goals, desires, or attitudes towards aspects of that situation. In philosophy, thinking about decision and action has been dominated by two models: Aristotle’s “practical syllogism” (APS), and the Bayesian calculus (BC), of which a particularly user-friendly form was elaborated by Richard Jeffrey (1965). Both start from more or less regimented conceptions of wants and beliefs, coming together to motivate and guide any intentional action. In the traditional picture represented by the APS, we start with a general apprehension of want or desirability as well as of the attendant circumstances. (BC), by contrast, starts with assessments of degrees of probability and degrees of desirability derived from preference rankings (Ramsey 1931).

Both models serve three very different and sometimes conflicting roles in discourse: (1) the articulation of first person decision-making; (2) third-person explanation of action; and (3) the provision of a tool for criticism of action, targeting practical irrationality. In the third role, both APS and BC proceed by detecting either a mistaken inference principle or an inconsistency among the premises included in different arguments simultaneously invoked. On a well-known analysis of the much discussed case of *akrasia*, for example, a comprehensive argument leads to the conclusion that it is best to do A, while a shorter argument, including a biased subset of the available considerations, results in the decision actually adopted, thus violating a “principle of continence” that requires that actions be based on the broadest available considerations (Davidson 1980). That analysis is not available to the Bayesian model, since by hypothesis that model takes into account the actual degrees of desirability and probability involved in bringing about the action. But BC can identify inconsistencies in the preference rankings implied by two different decisions: “if you cared *so much* – as indicated either by your professions of concern or by your previous decision – about X, why did you rank it so low in this other decision?”. In both models, the explanatory function may be at odds with the critical one. From the explanatory point of view, the action taken emerges out of the dynamics of whatever competing considerations have led to it. A charge of irrationality therefore competes with an alternative explanation which ascribes the failure of prediction to a misidentification of the agent’s beliefs and desires. Clear cases of irrationality can occur only when the subject’s explicit professions of belief and desire contradict one another. (de Sousa 1971, 2004).

That is just one way in which the explanatory mode may fail. An additional problem arises when we take account of all the available information about what a subject explicitly values. Since the statements that subjects are asked to rank in order to calibrate the desirability scales we ascribe to them include compound and conditional statements, the values we assign to those parameters may be distorted

by our well documented incapacity to make reliable inferences involving probability (Kahneman et al. 1982).

So while both models sometimes appear to fail of empirical adequacy by producing the wrong prediction, the critical perspective may simply regard these cases as manifesting the agent's irrationality. It is no defect in a critical tool that some practices deserve criticism. Logic also sometimes fails to represent the way people actually reason, but we don't take that as a sufficient reason to give up on the rules of logic.

We do, however, need to grant that both APS and BC, like logic itself, remain radically incomplete as accounts of how people behave. Consider first Aristotle's own classic example of a practical syllogism:

Every man should take walks,
I am a man,
(at once I take a walk.) (Nussbaum 1978, p. 40 (701a12–15)):

Obviously this is laughably unrealistic both as an explanation of why someone might take a walk and as an account of deliberation. A slightly more realistic story might go:

A walk would be good for me; but it's rainy and cold; besides, I have a lot of things to do. I can go to-morrow instead; anyway I have life insurance and no history of cardiovascular problems, and I've been walking quite a bit lately; besides I just don't really feel like it.

Yet even then, all of those considerations remain largely meaningless unless each can be quantified. A walk would be good: but *how* good? It's rainy and cold: but *how disagreeable* is that? *How urgent* are those other things I must do? And so on.

In sum, the APS has three major failings: First, it takes no account of degrees of belief or subjective probability: the belief component is treated as on/off. Second, it's not much better at degrees of desire. True, one could append a variable desirability measure to wants; but that wouldn't really help, in view of the third and particularly crippling problem, which is that the APS has no way of confronting and comparing different evaluative premises. There is no room in a practical syllogism for "on the other hand, I would prefer that other course of action."

Nevertheless, APS does have a major advantage over BC: it deals in explicit reasoning using language. I am inclined to think it describes *only* that kind of reasoning, although Aristotle himself appears to regard it as equally applicable to the "motions of animals" – the title of the book in which the example above is to be found. Animals share our interest in getting things right, but they do not share our explicit epistemic goals as such. *Truth, explanatory power, simplicity, and consistency* make literal sense only in connection with verbalized propositions. To have explicit beliefs is to be committed to rules of inference for categorical propositions, such as Modus Ponens, Modus Tollens, and conformity to mathematical theorems. Despite intriguing evidence that other mammals and birds are capable of some elementary arithmetic (Adessi et al. 2007), we do not expect the game of explicit formal reasoning to be played by non-human animals. (Non-human machines, by contrast – at least those equipped with "classical" or "von

Neumann” architecture – are better at formal inferences and calculations than we are. Computers are Second Track devices.)

It is also true, of course, that we do many things in much the same way as other mammals do them. These are among the behaviours typically controlled by “first track” processes. The brain uses a strictly Bayesian strategy in judging how best to hit a tennis ball, in the light of both visual input and prior expectation. (Körding and Wolpert 2004).

This example features all the essential features of agency. There are evaluative parameters (v) – values, or goals that might or might not be attained; and there are epistemic parameters (p) – beliefs or subjective probabilities. Both are subject to uncertainty, and both can be singly or jointly subject to inappropriate emotional interference. Furthermore, the tennis ball example illustrates the important point that uncertainty can pertain to different aspects of the situation: either to *prior expectations* or to *current sensory input*. Both modes of uncertainty are represented in the BC model, but not in APS. As was first expounded in (Levi 1967), there are at least two different ways in which we can think of “degrees of belief”. The standard way, going back to (Ramsey 1931), identifies it with subjective probability. But another important aspect of belief is its *stability*: the ease with which subjective probability might be modified by new evidence. To illustrate the difference between subjective probability and stability, suppose I toss an unsuspected but normal-seeming coin. You will typically think it fair to make an even bet on either Heads or Tails, indicating that you attribute a probability of one half both to its landing on Heads and to its landing on Tails on any one toss. But that expectation might be disrupted if in the first ten tosses you get a run of 10 consecutive heads: in the light of that result, you may now judge it less likely that the coin is fair, and change your probability assignment accordingly. By contrast, if you have already watched two thousand tosses, yielding 972 Heads to 1,028 Tails, a run of ten consecutive heads will not affect your assessment of the coin’s fairness.

When calculations of probability are explicit, we have systematic ways of making calculations but we often get them wrong. By contrast, we are quite sensitive to differences in frequencies among actual outcomes. (Whitlow and Estes 1979). The difference language makes to second-track processes rests not on the capacity for verbal communication, but on those extensions of that capacity that stem from Aristotle’s discovery of *logical form*. Aristotle was the first, at least in the Western tradition, to identify forms of inference independent of their content. On that simple fact the entire field of computer science depends: since computers know nothing, they could do nothing if reasoning depended on understanding. The obverse of the irrelevance of content to validity is that the scope of discourse is universal. Eyes see only sights; ears hear only sounds. But precisely in virtue of its essential abstraction from the input of specific transducers, language as such can in principle be about anything. Among other consequences, this enables information from one modality to be conveyed to others (Carruthers 2002). When problems are both novel and complicated, this is particularly crucial to the elaboration of responses that go beyond those programmed into the intuitive track.

The principal virtues of the BC model stem, as we saw, from the fact that the model works with *degrees* of belief and desire. Their interaction is represented as a dynamic interaction of vectors, and, as we saw in the case of K rding's tennis player, it takes account of real-time adjustments of behaviour in light of the interaction of current evidential data and prior expectations. It works for non human as well as human animals; but in humans its role, to be realistic, must be regarded as explanatory rather than critical. The reason is that a criticism can be legitimate only where verbal confirmation of an observer's ascriptions of beliefs and desires can be obtained: otherwise, there is always an alternative interpretation available, on which what looks like inconsistency is really a change of mind or else is due to mistakes in the original assignment of values to the v and p parameters. And while humans can provide that kind of corroboration in general, agents' quantitative assessment of their own degrees of confidence or of desire are notoriously unreliable. On the "two track" perspective, this is to be expected, since we have no conscious awareness of the processes that underlie our intuitive decisions. As evidenced by a growing body of data, subjects, like observers, have only inferential access to the mental processes that determine decisions taken by the intuitive track (Wilson 2002).

Furthermore, BC is also incomplete or simplistic in other ways, some of which stem from the attempt to apply it explicitly. Sometimes values will be practically incommensurable within a broad range (de Sousa 1974). At other times a mathematical equivalence will give rise to different subjective assessment dependent on framing and formulation effects. The examples are familiar (Tversky and Kahneman 1981): subjects strongly prefer a policy resulting in 80% survival to one involving 20% deaths. And the death of 50 passengers in separate auto accidents is judged much less catastrophic than the death of 50 in a single plane crash. A striking effect of the tendency to concentrate on the size of a given disaster and ignore greater but less salient dangers is this: in the year following the 9/11 attacks, almost as many additional deaths as those directly caused by the terrorist attacks were due to the additional (and far more risky²) miles traveled by car in response to the fear (and perhaps also added inconvenience) of air travel. (Blalock et al. 2005).

The sources of these anomalies in our assessment of risk have been extensively discussed.³ But one very general reason deserves to be stressed: We're bad at reasoning explicitly about situations that do not trigger appropriate first track responses. To get things right when we are confronted with complex situations, we need language, math, and logic. But we are still strongly, and sometimes disastrously, inclined to bypass those tools and trust our emotions.

²Depending on how it is computed, flying in a commercial airliner is about an order of magnitude less dangerous than riding in a car. One source cites the rate of deaths per million passenger miles at 0.03 in certified airline carriers compared to about 2 per million car-occupant passenger mile, which makes cars about 7 times more likely to kill you than commercial planes (Dever and Champagne 1984, p. 362). A more recent statistic is that the risk of dying is about the same, per passenger *hour* in plane or car. Assuming that the average speed of an airliner is at least ten times the average speed of a car, this yields a somewhat higher ratio but one of the same magnitude. (Levitt and Dubner 2005, p. 151).

³Some classic sources are Kahneman et al. (1982), and Slovic (2000).

3 The Circle of Emotional Appraisal

Even in our attempts to reason rigorously, we are susceptible to the influence of emotions. Nico Frijda has identified a number of promising hypotheses about how the “Laws of emotion” might differ from the laws of logic. (Frijda 2007) His “Law of Apparent Reality”, for example, involves “visual presence, temporal imminence, earlier bodily encounters, pain” (Frijda 2007, p. 10), all of which are irrelevant to the truth of a simply logical or inductive inference. Another emotional processes that doesn’t conform to what cool common-sense would expect is *hyperbolic discounting* of future prospects (Ainslie 1992, 2001), which seems arbitrary in preference to a more linear formula. A third concerns our assessments of the past: common-sense suggests that our assessment of lived episodes should reflect some computation of the pleasure afforded by each period weighted by its duration. In fact, however, the *Peak-End Principle* we intuitively use to evaluate past episodes defies this rule of common sense, discarding from the calculation all but the extreme and the final components of a complex episode. (Kahneman et al. 1993).

Friends of the Intuitive Track have stressed the virtues of intuitive and emotional responses. Emotions program “fast and frugal” scripts that efficiently bypass excessive calculations (Gigerenzer et al. 1999). But the further away our lives get from that of our speechless ancestors – the more technology is essentially involved – the more we confront problems for which our intuitive resources have not prepared us. Getting to Mars is not something we can do by trusting atavistic intuitions. We need calculation, explicit logic and mathematics, and the computers that are at long last speeding up the arduous processes of calculation to match those of intuitive processing.

That does not mean, however, that we can sideline the role of emotions. In relation to the mind’s two tracks, emotions are intrinsically hybrid: as intentional states, they commonly have articulable objects about which we can reason explicitly. But as bodily states involving complex action-readiness (Frijda 2007) their scripts are only partly within the control of the analytic system. So they belong to both the Intuitive and the Analytic systems. That doesn’t necessarily mean that they combine the virtues of both: on the contrary, it means they should remain suspect to either point of view.

Emotions also bridge thought and action, notably in the specific sense that they are involved in both strategic and epistemic rationality. The distinction is an important one, but it is not exhaustive. Both kinds of rationality are assessed in terms of the likelihood of success of their respective aims. Strategic rationality relates to a specific goal, and its measure is the likelihood of success in reaching that goal. Epistemic rationality is assessed by reference to a limited subset of possible goals, namely the epistemic goals mentioned above, and more specifically by the likelihood that the process of acquiring a belief employed in a particular case will lead to epistemic success. The relation between practical and epistemic rationality has long been a matter of dispute. In one perspective going back to Socrates, practice presupposes truth, and “virtue is knowledge” (Plato 1997). It is also exemplified by William Clifford’s prescription for the ethics of belief: “it is wrong always,

everywhere, and for anyone, to believe anything upon insufficient evidence.” (Clifford 1886). A contrary tradition goes back to Protagoras, who professes to be unconcerned with truth but only with practical effectiveness, and it is exemplified by one variant of philosophy of pragmatism, in William James’s (1979) response to Clifford.

The debate leads to a stalemate (de Sousa 2003). In the fundamental *value-belief-means-end* nexus, epistemic and practical rationality can clash. When they do, each can make a case for subsuming the other; but neither can get beyond begging the question. One can hear principled outrage on both sides: Should one not care more about truth than advantage? (and your practical rationality be damned), say Socrates and Clifford. But Protagoras and James respond: Practice subsumes truth: Should one not care about real consequences and not abstract truth? (and your epistemic scruples be damned). Only a third form of rationality can adjudicate without begging the question, namely one capable of judging the “appropriateness” of different *kinds of appropriateness*. Call that type of rationality *axiological*, because emotions function as perceptions of value. Epistemic feeling – such as doubt, certainty, the feeling of rightness, the feeling of knowing – are called on to arbitrate (de Sousa 2008). The stance one chooses to take towards Pascal’s notorious wager, for example, is inevitably determined by one’s emotional response to the question of whether it is appropriate to judge religious belief on purely epistemic criteria or on the contrary to regard it as a practical problem.⁴ Emotions, then, are both judge and party. Such is the circle of emotional validation. Not all circles are vicious. If a circle is large and inclusive enough, it gets rehabilitated as a coherence account of justified belief. This is reflective equilibrium.

Reflective equilibrium cannot evade the crucial role played by emotions. Emotions quite properly affect goals and values. Indeed, if there were no emotions, it is debatable whether we could intelligibly speak of values at all (Prinz 2007). But one can still worry about when and how the influence of emotions is legitimate and when it is not. One can have doubts, for example, when they affect beliefs directly, in the way just alluded to, by legitimizing a strategic rather than an epistemic appraisal of belief. Furthermore, emotions can apparently affect the belief-desire complex directly, without passing through a detectable prior process of affecting the one or the other. Emotional attitudes apply to meta-cognitive judgments of appropriateness where the rationality or reasonableness of emotions themselves are in question.

Before I elaborate on this, consider an example. Should we fear death? Lucretius, drawing on Epicurus, argued that fear of death is irrational, on the ground that I can never experience the harm of death. I can’t feel the harm of death while still alive, since I’m not dead; and I won’t feel it when I am dead, because then I will feel nothing (Lucretius 1951: Bk.III, 830–840).

⁴Pascal argues that even if we assume the probability of God’s existence to be arbitrarily small, the infinite expected value of the stakes involved (eternal heaven or eternal hell) nullify the epistemic disadvantage of belief and make it the preferable option (Pascal 1951, §233).

But now if the thought that I will feel nothing at a future time is consoling for me now, then some future facts matter to me now. And if that is so, then – as Philip Larkin pleads – why shouldn't that very thought distress rather than console me? "And specious stuff that says No rational being/Can fear a thing it will not feel, not seeing/That this *is* what we fear. . . ." (Larkin 1977, my emphasis). The Epicurean argument does not always dissolve the fear of death; yet sometimes it does: I, for one, do find the argument compelling on its own terms. The moral is that in some cases, only *one's emotional attitude itself determines what emotional response is rational*.

More generally, "You should (or shouldn't) care" can be effectively justified to any particular person only by appealing to what already concerns them. In the final analysis, the normative claims of rationality can be justified only by appeal to certain specific emotions. Both moral and epistemic feelings act as arbiters of rightness. But unfortunately there is no compelling reason to expect all of our biologically evolved emotional capacities to serve our present purposes, or even to be mutually coherent.

4 Relative Rationality

How then are we to characterize rationality? In assessing an inference, only a feeling of rightness can determine whether p & $(p \rightarrow q)$ should compel us to believe q , or to reject *either* p or $(p \rightarrow q)$. That feeling of rightness – in a *reasonable* person, a qualification which evidently invites a reduplication of the problem – will emerge out of a large number of relevant considerations about the context of the argument, as well as any independent inclinations to believe the premises or to disbelieve the conclusion. Similarly, in the case of a moral problem, we typically weigh the undesirability of consequences against the desirability of "principle", looking at each in the light of the other. As in the case of factual or logical inferences, reflective equilibrium affords the only prospect of resolution. And *what needs to be placed in equilibrium are emotions*.

The "Trolley Problem" provides an illustration. A brief reminder of this now well-known thought experiment should suffice. A trolley has lost its brakes and is heading down a line on which, if it proceeds unheeded, it will inevitably kill five workers. In *Scenario I*, you are in a position to flip a switch, diverting the trolley onto another track, where it will, with equal certainty, kill one lone worker. In *Scenario II*, you are on a bridge overlooking the track; there is no switch, but you could push a large man from the bridge onto the track. He will certainly be killed, but the trolley's progress will be blocked, saving the five on the track. In terms of the consequentialist calculus based on the value of saving lives, the two situations are equivalent. Yet while most people respond that they would flip the switch in the first scenario, most say they would not push the fat man onto the track in the second (Greene 2008).

One interpretation of these results is that the different responses to scenarios I and II are due to the degree of personal involvement in the causation of the event. In Scenario II, the involvement of the agent is more "personal", and the discrepancy looks like the difference between the difficulty of killing someone in hand-to-hand

combat compared with launching a bomb or rocket at a distance. Whatever the exact mechanisms may be that result in these differential responses, they appear to be so ingrained that it takes brain damage to undo the effect:

Six patients with focal bilateral damage to the ventromedial prefrontal cortex (VMPC), a brain region necessary for the normal generation of emotions and, in particular, social emotions, produce an abnormally 'utilitarian' pattern of judgments on moral dilemmas that pit compelling considerations of aggregate welfare against highly emotionally aversive behaviours (Koenigs et al. 2007).

From this, it would be hasty to infer that utilitarians have defective brains. We know all too well that intact brains don't infallibly arrive at the right moral judgments; and the mere fact that many people agree on a moral judgment no more warrants its correctness than the popularity of McDonald's food proves it to be healthy. What the case does illustrate is that our emotional responses deliver contextually relative assessments of rationality.

A judgment of rationality can be contextually relative in at least two senses. First, it can arise in the light of principles that are *more or less obligatory*. Second, it can be grounded (and it can seem *reasonable* for it to be grounded) in a more or less inclusive *framework*.

4.1 Rationality, Obligatory and Optional

Some principles of inference are incontrovertible. Their validity in ordinary reasoning is unquestionable, even if someone fails to acknowledge it. Modus Ponens, Modus Tollens, the law of non-contradiction, and the rules of elementary arithmetic are, in this sense, *compulsory*. This does not mean, however, that we can provide a *proof* of their validity. On the contrary: what makes argument about such basic principles particularly frustrating is that they are "self-evident", which means that any argument for them tends to make them seem less rather than more compelling. As Lewis Carroll's puzzle of Achilles and the Tortoise shows, the provision of a "proof" – i.e. of an explicit premise from which it follows deductively that Modus Ponens is correct – generates an infinite regress (Carroll 1895). Such principles, or better practices, need to be innate, in order to carry the conviction on which they rely. Although it is sometimes difficult to make it clear to subjects that they are asked to perform Modus Ponens, it cannot be extensively violated without a disintegration of rational discourse.⁵

Other principles of inference might be said to be *weakly compulsory*, in the sense that it is indeed possible to *demonstrate* that they are correct, but that doesn't mean

⁵A nice but fictional illustration of the disintegration of discourse that results from ignoring elementary rules of logic is in one of Douglas Hofstadter's charming elaborations on Achilles and the Tortoise (Hofstadter 1980, pp. 177–180). For the difficulty of getting subjects to confine themselves to the terms of a deductive argument, see (Luriiia 1976). Not everyone agrees that contradictions have catastrophic consequences for rational discourse. Peng and Nisbett (1999) have claimed to find educated Chinese subjects who don't object to believing contradictions, and Graham Priest (1997) has argued that in the right context, the proliferation of inferences derivable from a contradiction is effectively contained.

it's always possible to persuade an otherwise rational person. A nice example of this is provided by the controversy raised by the problem known as the "Monty Hall problem":

Three doors are visible, and you know that behind one of them stands a Cadillac, while each of the two others hides a goat. I ask you to guess which door is the good one. I then open one of the other doors, revealing a goat. Now I ask you to bet on which of the two remaining closed doors is the good one: the one you originally picked, or the other one? It is tempting to reason: since there are just two doors, it makes no difference. You could switch or stay at random. In fact, however, you stand to win two thirds of the time if you switch; while if you stay with your original choice, you will lose two thirds of the time. For of all the times you start playing this game, pointing at random will pick the Cadillac door only once in three.⁶

In this and many other cases familiar from (Kahneman et al. 1982), our intuitive answers are often objectively wrong. It doesn't follow, needless to say, that "evolution failed us", since it is plausible to speculate that under the constraints likely to be in effect during the environment of evolutionary adaptation (EEA), the decision procedure in question might have been the best available.

In these compulsory cases, we might expect that once the problem is sufficiently well defined, we can give conclusive reasons for the superiority of one argument or method over another. This class of examples differ from the "strongly compulsory" ones in that they make no claim to foundational status. As a result, they admit of (conclusive) justification. Anyone inclined to dispute the standard solution to the Monty Hall problem can be invited to put their money behind their principle.

In other cases, however, and particularly where the reasonableness of emotional responses are themselves in question, there may be two conflicting and equally compelling answers. We are left with real paradox. We've already seen three examples of this: the Epicurus argument against fear of death; the Peak-End principle, and hyperbolic discounting. The last two, unlike the other, appear to be both *surprising* and *universal*, which seems surprising in itself. But in those examples the arguments themselves didn't carry conviction on logical grounds alone. In a particularly puzzling class of cases, the conflicting arguments have the logical force of a classic antinomy. Such is Newcomb's problem, in which a dominance argument on one side and a Bayesian reasoning on the other seem equally impregnable, though their conclusions are radically incompatible. (Nozick 1970).⁷

⁶This puzzle had been around for some years before becoming widely known as the Monty Hall problem. Hundreds of mathematicians and statisticians, it was reported, got it wrong (Martin 1992, p. 43).

⁷You may take one or both of two boxes. One is transparent and contains €1000. What the second, opaque box contains depends on what a hitherto apparently infallible predictor has predicted you will do. If he thought you would take just the opaque box, that box contains €1 million; if he thought you would take both, it is empty. The *Bayesian* argument supports taking just one box, given the high probability that the predictor got it right. The *dominance* argument supports taking both, since the content of the box is already determined and is strictly causally independent of the present choice.

4.2 Context and Framing

The other way that our assessments can be contextually relative relates to the breadth of the frame in which it is placed. Andrea Yates drowned her five children, in obedience, she said, to the voice of God. In her first trial, the insanity defense was not admitted, in view of the methodical way in which she proceeded. Yet should not the project itself of drowning your five children be deemed irrational? Not necessarily: for consider the case of Abraham, or that of Agamemnon, both of whom agreed to slaughter their child in obedience to a deity. In that context, neither is irrational. Yet again, is that context itself not profoundly irrational? There is not in general an objective, absolute context in which the question can always be conclusively answered.

5 Fear as a Measure of Risk

A natural, common sense hypothesis is that the biological function of fear is as a *measure of risk*. If that is right, we might expect that varieties of fear – or the way they work – would reflect the ambiguity noted in Section 1 above. This would show up as follows in terms of the standard formula expressing expected utility,

$$V = \sum_{i=1}^n (p_i \times v_i) :$$

fear can affect the result V in several ways, such as by affecting p directly, by affecting v , or by somehow short-circuiting both to influence the result without affecting either of the input variables. It isn't easy to see just how we could tell which is going on in any particular case. But it is clear that in many cases fear is very far from tracking risk in the sense of overall expected utility. An example:

In the five years from September 2001 to September 2006, about 3,500 people have been killed by terrorists. During the same period, very roughly 200,000 have been victims of fatal road accidents. It's been estimated that about the same number have been killed by guns, and there have been about as many iatrogenic deaths as both the last put together (Feckler 2005)⁸, for a total of 800,000 people. It follows that an American is well over 200 times more likely to die of guns, traffic accidents, or medical errors than of terrorist attacks. In a study by the Federal Reserve Bank of New York, it's been estimated that in a comparable period the increase in expenditure devoted to Homeland Security in response to the terrorist attacks has amounted to about a quarter of 1% of GDP (Hobijn and Sager 2007). It follows that if proportional resources were to be devoted to prevention of those non-terrorist sources of danger, that would take up 50% of American GDP.

⁸This statistic is arguably suspect in motivation, since it is provided by an avowed partisan of "one man one gun", but I have no reason to doubt its correctness.

The relevance of this example admittedly rests on a rather large assumption, which is that public policies are to some extent determined by perceived fear in the public. It may be slightly more plausible to attribute such policies to the politicians' fear of not getting re-elected. They will then fall into place alongside other idiocies of public policy, such as the "war on drugs", or the reliance on coal-burning plants rather than nuclear power for generating electricity.⁹

More direct evidence exists that a global assessment of a non-specific "risk" can be affected by factors linked only indirectly or not at all to the probability of an event. Accepted levels of risk in voluntary activities is proportional to the 3rd power of benefit for that activity. (Starr 1969). Level of risk accepted for voluntary activities (skiing, or skydiving,) is about 1,000 times the level accepted for involuntary activities.

6 Effects of Metacognition

Although it has been long established that some of the strongest "basic" emotions can be evoked in the absence of any cognitive awareness (Zajonc 2000), it is equally well known that the character and valence of emotions, including pleasure and pain, can be radically affected by beliefs or attitudes. In particular, some emotions, including fear and pleasure, can take instances of themselves as objects. This can work to enhance a pleasant emotion, to mitigate an unpleasant one, or even to reverse its valence altogether. In some cases, fear is actually experienced as pleasurable or as an enhancement of pleasure. These are cases where there is a metacognitive frame around the experience that amounts to a conviction that any actual danger is absent or minimal (as in horror movies or fairground rides). There are also cases where the intrinsic quality of fear is held to spice up the pursuit of some thrill. In those cases, then, the unpleasantness of the danger posited as the object of fear is mitigated by the intrinsic pleasantness of the emotion. Generally speaking, however, fear is intrinsically unpleasant; in that case, the intrinsic disutility of fear must be added to the disutility of what is feared. The first consequence of this is that the intrinsic disvalue of fear must be added to the prospect feared. The Bayesian formula becomes recursive, as fear of fear itself increases the present fear:

$$V(\text{fear at } t + 1) = \sum_{i=1}^n (p_{i(at\ t)} \times v_{i(at\ t)}) + V(\text{fear at } t)$$

One can see how this formula might represent a panic that feeds on itself, in such a way as to outstrip the usefulness of its biological signaling function.

⁹Economically viable levels of safety for nuclear power (as well as experience over half a century) point to a risk of death some forty to a hundred times lower than that now associated with coal (Starr 1969, p. 1237). These figures ignore other drawbacks of coal generated power, such as pollution and greenhouse gas production. They also ignore other objections to nuclear power, based on technological problems such as the disposal of waste and political ones based on the higher cost of security. My thanks to the Editors for pointing this out.

As a measure of risk, fear should affect just p or v in the Bayesian formula, but not both. Becker and Rubinstein have argued, however, that fear can irrationally affect both at once:

[A]n exogenous shock to the underlying probabilities affects agents' choices via two different channels: (i) the risk channel: a change in the underlying *probabilities* keeping (marginal) utility in each state constant; (ii) the fear channel: a change in the underlying probabilities also determines agents' optimal choice by affecting the *expected utility* from consumption in each state (Becker and Rubinstein 2004. My emphasis, to mitigate the difference in terminology).

Here is one specific way that they argue p and v get confounded. Citing an analysis of the effect of terrorist attacks on business-cycles in the Israeli economy (Eckstein and Tsiddon 2003), Becker and Rubinstein point out that when terror endangers people's lives, their estimation of the value of the future relative to the present is reduced. As a result, investment declines, as do long-run incomes. A very low increase in the probability of death due to terror nonetheless generates a large effect, by modifying the value placed on the outcome.

Some more general distortions in the perception of risk have been explored by (Fischhoff et al. 1978), who have shown that when risk levels are deemed more or less acceptable, there is a confounding of estimates of benefit with estimates of acceptable risk: in other words, if you think a process is beneficial, you will think it safe enough; conversely, if you think it is not safe, you will forget about the benefits as well. Obviously, from a perspective of broadly Bayesian rationality, this confusion is not a good thing.

7 Application to Risky Technology

The exponential progress of technology in the past century (Kurzweil 2005) affords a particularly tempting opportunity for assessing the consequences of our emotional responses. Three domains of technology provide particularly good illustrations of some of the issues involved: nuclear power, genetic modification of foodstuffs, and nanotechnology.

I argued in Section 5 above that when judged in relation to the real dangers and documented fatalities attributable to coal mining and use, resistance to nuclear power can seem entirely irrational. The sort of considerations just alluded to can help to explain why the attitudes in question are so tenacious: If a nuclear accident has a tiny but real probability, the value of the future is reduced: so when we compute the desirability of the outcome, we don't just apply the Bayesian formula to *life as we know it* and *life after a nuclear accident*. The very possibility of a nuclear accident affects our estimate of the value of *life as we know it*. There is a kind of double counting here: it's not just that a future with nuclear waste is less valuable as well as more probable given the existence of one more nuclear plant. Rather it's that the building of the nuclear plant reduces the value of life *even if no accident ever occurs*, simply by making its mere possibility more vivid. Is such double counting

irrational? It might be viewed as a rational form of the “social construction” of risk, or it might be looked at as one more way in which the emotional processing of risk leads to irrational assessments.

Similarly, the negative feelings generated – particularly in Europe – by genetically modified agricultural products seems to be based on a number of different factors, including political objections to the privatizing of biological organisms, and the perceived threat to biodiversity. But much of it appears to be driven by a visceral response to processes and products felt to be “unnatural”. (Anonymous 2006). The cogency of that response, however, cannot stand critical scrutiny, since it is evident that “naturalness” is not a sufficient condition of goodness even for the most enthusiastic environmentalists, who are unlikely to have qualms about doing away with “natural” organisms such as the smallpox virus or the syphilis bacterium, although both of those are among endangered natural organisms. At the very least, emotional responses must be scrutinized for inconsistencies that will make it clear that we aren’t really concerned with the “naturalness” of an organism, but with entirely different issues masked by that slogan.

The case of nanotechnology is somewhat different again, because unlike nuclear power and genetic manipulation of organisms, it has yet to yield any actual results or indeed coalesce into a single recognizable field. Just as major technological inventions are by definition unpredictable (for if they had been predicted, they would not be new inventions), so their costs and benefits, and the probabilities of those costs and benefits, are almost equally impossible to assess in advance. In the face of truly radical uncertainty, a Bayesian calculation can’t get going simply because we don’t know how to assign values to the relevant parameters. The resulting situation can be described in one of two ways. The first way is to insist that since what enters into a Bayesian formula are *subjective* probabilities, the fact that no grounds can be found for the assignment is of no consequence. Estimates of both the probability and the value of various outcomes can be made arbitrarily. The second way is to ignore both probabilities and the value of outcomes, and to invoke the blanket “fire-wall” of a “precautionary principle” to reject technological change. Actually these two approaches, although they are rhetorically distinct, could turn out to be equivalent in their consequences, depending on the assignments made in the Bayesian formula.¹⁰

Either way, it is clear that nanotechnology, to a greater extent than the others mentioned here, gives rise to what has become known as the “Collingridge dilemma”: before a technology gets underway, we could monitor and control it, but we lack the knowledge of its consequences that would be required in order to do so intelligently. Once that information exists, however, the technology will be entrenched

¹⁰As the Editors helpfully point out, there seems to be an obvious alternative, which is to carry on more research until it can be established that a technology is safe. But as the discussion in the next paragraphs suggests, some proposed application of the precautionary principle apply to domains where the large-scale research that alone can certify safety requires that large numbers of subjects be involved, and so be put at risk. Conversely, while a proposed technology is withheld until it is deemed “sufficiently” safe, lives may be lost owing to its unavailability.

and it will be extremely difficult to modify or control it (Collingridge 1980). It is in cases like this that the Precautionary Principle may have some appeal: radical uncertainty about a particular domain could seem to warrant blind resistance to its exploration. On the other hand, such blanket rejection looks irrational in the light of the history of benefits from technology as well as the poor track record of the predictions of disaster that have attended most new technologies.¹¹ And in any case, while the Precautionary Principle may well be the only available tool specifically tailored to that degree of ignorance, that is not reason enough to recommend it. For as Cass Sunstein (2005) has forcefully argued, it undermines itself. By the very same reasoning as might be used to argue that nanotechnology (or any other radically new technological venture) poses unknown dangers, and should therefore not be undertaken, it can be countered that it might present unknown benefits that would protect us against more serious dangers, and that it must therefore be explored.

Furthermore, there is some additional reason to believe that the appeal of the precautionary principle is due to a primitive mechanism that belongs to first track processing, and that kicks in without calculation or explicit endorsement by second track reasoning in the face of “unknown unknowns.” Such a mechanism has been hypothesized to lie at the heart of both religious and social rites, as well as causing the pathological rituals associated with obsessive compulsive disorder (OCD) (Boyer and Liénard 2006). Neither association recommends it. And in the light of that hypothesis, it is not surprising that attitudes to the risks of nanotechnology appear to be governed by a kind of infantile logic that resembles a child’s “I won’t taste it because I don’t like it”. There is evidence that attitudes to this technology are strongly correlated with epistemically irrelevant factors such as race, gender, political ideology, and political attitudes. Information acquired tends merely to reinforce attitudes predictable on the basis of ideology, rather than affecting beliefs in accordance with its evidential status (Kahan et al. 2007, 2008).

8 Conclusion: Advice to Philosopher-Kings

First track processes are obviously not selected to deal with the kind of problems that arise from the risks and benefits of advanced technology. It is therefore to be expected that our intuitions and emotional responses in this area will not be particularly reliable guides to policy. The experiments cited in the last section are particularly disconcerting, since they suggest that epistemic rationality plays no role

¹¹“It was claimed that trains would blight crops with their smoke and terrify livestock with their noise, that people would asphyxiate if carried at speeds of more than twenty miles per hour, and that hundreds would yearly die beneath locomotive wheels or in fires and boiler explosions. Many saw the railway as a threat to the social order, allowing the lower classes to travel too freely, weakening moral standards and dissolving the traditional bonds of community; John Ruskin, campaigning to exclude railways from the Lake District, warned in 1875 of ‘the certainty. . . of the deterioration of moral character in the inhabitants of every district penetrated by the railway’.” (Harrington 1994, p. 15).

at all in the elaboration of attitudes to nanotechnology. In the other cases I have considered, however, it seems we can sum up the types of role played by first-track emotional response – at the price of only minimal simplification – as involving one of more of four mechanisms that bring some sort of systematic distortion to the Bayesian decision process:

- (1) Emotions affect (or constitute) a change in the value of the belief parameter p .
- (2) Emotions affect (or constitute) a change in the desirability parameter v .
- (3) Emotions somehow effect an immediate apprehension of “risk” as if there had been a kind of merger of p and v into a blended value that both contradicts the acknowledged values of p and v and resists decomposition into separate parameters.
- (4) Emotions driven by temperament or ideology can somehow short-circuit an estimate of expected value altogether by effecting a non-Bayesian (on/off) input directly into the conclusion.

Emotions, and particularly fear, are subject to bootstrapping effects: since they are essential arbitrators of value, as argued in Section 3 above, they can’t be merely regimented in the light of values independently assessed. I have argued that confounding the parameters in the complex conception of risk can cause runaway positive feedback effects, double counting, and in other ways illegitimately change belief on the basis of epistemically irrelevant factors. It is facile, if not fatuous, to conclude that we should manipulate emotion in benevolent ways. The difficult question raised by that conclusion is who “we” are to do anything of the sort. In any case, emotion itself determines the values in the name of which we act: what I have called the circle of emotional appraisal leaves us with no entirely independent objective point of view from which to decide what to do.

What we can do, as scholars or philosophers, is articulate as clearly as possible the reasons for distrusting our emotions, even as we appeal to some of our emotions, including epistemic feelings of doubt, of “rightness”, or of relative certainty. It can be helpful, in particular, to distinguish three phases in the process leading to any decision concerning a major issue of policy: (A) *Discovery* (of relevant facts and preferences or values); (B) *Justification* (of the judgments discovered, and inferences made from them), and (C) *Motivation* (of the “detachment” of judgment in action). Emotions are involved in phases (A) and (C). In (A), they provide prima facie evidence of caring or concern (Roberts 1988): what we notice is a sound prima facie indicator of what matters to us. And in (C), emotions are crucial because only what we care about is capable of motivating action. But in stage (B), the all-important intermediate stage of justification, we need the solid, language-based intellectual nitty-gritty of explicit argument, good statistics, measurements of probabilities and outcome values, stripped of the power of rituals or immediate emotional response.

If scholars and philosophers were elected to the role of Philosopher-Kings and could act as the Providential State, they could not altogether escape the obligation to manipulate the emotions of the public at stage (A) and, once the work of justification

at stage (B) is done, at stage (C). This could be done in the spirit of Sunstein and Thaler (2003)'s policy of "libertarian paternalism". But at least one can hope that it might be done with maximal transparency. For as Doris Lessing (1987) has pointed out, there is hope that people's freedom can be enhanced by making them aware of the emotional forces to which their nature as humans beings subjects them. Insofar as awareness of the risk of manipulation may lead to greater autonomy, it can guide a self-conscious policy of benevolent manipulation.

References

- Addressi, E., L., Crescimbene, and E. E., Visalberghi. 2007. Do capuchin monkeys (*Cebus apella*) use tokens as symbols? *Proceedings of the Royal Society B* 274: 2579–2585.
- Ainslie, G. 1992. *Picoeconomics: The strategic Interaction of Successive Motivational States within the Person*. Cambridge: Cambridge University Press.
- Ainslie, G. 2001. *Breakdown of Will*. Cambridge: Cambridge University Press.
- Anonymous. 2006. Voting with your trolley: Can you really change the world just by buying certain foods? *The Economist*, 2007 December, accessed online 2009/03/09
- Becker, G., and Y., Rubinstein. 2004. *Fear and the Response to Terrorism: An Economic Analysis*. Online at <http://www.ilr.cornell.edu/international/events /upload/ BeckerrubinsteinPaper.pdf> (Accessed 2009/02/20)
- Blalock, G., V., Kadiyali, and D., Simon. 2005. *The Impact of 9/11 on Road Fatalities: The Other Lives Lost to Terrorism*.
- Bora, A. 2007. Risk, risk society, risk behavior, and social problems. In *Blackwell Encyclopedia of Sociology*. G. Ritzer, ed., Blackwell Reference Online, accessed 2009/03/09. Oxford: Blackwell.
- Boyer, P., and P., Liénard. 2006. Why ritualized behavior? Precaution systems and action parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences* 29: 1–56.
- Carroll, L. 1895. What the tortoise said to achilles. *Mind* 4: 278–280. Online at <http://www.ditext.com/carroll/tortoise.html>
- Carruthers, P. 2002. The cognitive functions of language. *Behavioral and Brain Sciences* 25(6): 657–674.
- Clifford, W. K. 1886. The ethics of belief. In *Lectures and Essays (2nd Ed.)*. L. Stephen and F. Pollock, eds., London: Macmillan.
- Collingridge, D. 1980. *The Social Control of Technology*. New York: St Martin's Press.
- Davidson, D. 1980. How is weakness of the will possible? In *Essays on Actions and Events*, 21–43, Oxford: Oxford University Press, Clarendon.
- de Sousa, R. 1971. How to give a piece of your mind, or the logic of belief and assent. *Review of Metaphysics* 25: 51–79.
- de Sousa, R. 1974. The good and the true. *Mind* 83: 534–551.
- de Sousa, R. 2003. Paradoxical emotions. In *Weakness of Will and Practical Irrationality*. S. Stroud and C. Tappolet, eds., 274–297. Oxford; New York: Oxford University Press.
- de Sousa, R. 2004. Rational animals: What the bravest lion won't risk. *Croatian Journal of Philosophy* 4(12): 365–386.
- de Sousa, R. 2008. Epistemic feelings. In *Epistemology and Emotions*. G. Brun, U. Doguoglu, and D. Kuenzle, eds., 185–204. Aldershot: Ashgate.
- Dever, G. A., and F., Champagne. 1984. *Epidemiology in Health Services Management*. Sudbury, MA: Jones & Bartlett Publishers.
- Eckstein, Z., and D., Tsiddon. 2003. *Macroeconomic Consequences of Terror: Theory and the Case of Israel*. Unpublished manuscript.
- Feckler, M. L. 2005. "Firearms in America: The Facts," Newsmax.Com, 10 August.

- Fischhoff, B. et al.. 1978. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences* 9(2): 927–952.
- Frijda, N. 2007. *The Laws of Emotion*. Hove: Erlbaum.
- Gigerenzer, G., and P., Todd, and ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. New York: Oxford University Press.
- Greene, J. D. 2008. The secret joke of Kant's soul. In *Moral Psychology, Vol. 3: The Neuroscience of Morality*. W. Sinnott-Armstrong, ed., 35–81. Cambridge, MA: MIT Press.
- Harrington, R. 1994. The neuroses of the railway. *History Today* 44 (July): 15–21.
- Hobijn, B., and E., Sager. 2007. What has homeland security cost? An assessment 2001–2005. *FBNY Current Issues in Economics and Finance* 13(2), February: 1–7.
- Hofstadter, D. R. 1980. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Random House, New York.
- James, W. 1979. The will to believe. In *The Will to Believe: And Other Essay in Popular Philosophy*. F. H. Burkhardt, ed., Cambridge, MA: Harvard University Press.
- Jeffrey, R. C. 1965. *The Logic of Decision*. New York: McGraw Hill.
- Kahan, D. M. et al. 2007. *Nanotechnology Risk Perceptions: The Influence of Affect and Values*. Project on emergent nanotechnologies, 18. Woodrow Wilson International Center for Scholars.
- Kahan, D. M. et al. 2008. *The Future of Nanotechnology Risk Perceptions: An Experimental Investigation of Two Hypotheses*. Harvard Law School Program on Risk Regulation Research Paper, No. 08-24. Cambridge. Available at SSRN: <http://ssrn.com/abstract=1089230>.
- Kahneman, D., B., Fredrickson, C., Schreiber, and D., Redelmeier. 1993. When more pain is preferred to less: Adding a better end. *Psychological Science* 4: 401–405.
- Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge and New York: Cambridge University Press.
- Koenigs, M., L., Young, R., Adolphs, D., Tranel, F., Cushman, M., Hauser, and A., Damasio. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446(7138): 908–911.
- Kurzweil, R. 2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Penguin, Viking.
- Körding, K., and D. M., Wolpert. 2004. Bayesian integration in sensorimotor learning. *Nature* 427(6971), 1915 January: 244–247.
- Larkin, P. 1977. Aubade. *Times Literary Supplement*, 1923, December.
- Lessing, D. 1987. *Prisons we Choose to Live Inside*. New York: Harper Collins.
- Levi, I. 1967. *Gambling with Truth*. New York: Alfred A. Knopf.
- Levitt, S. D., and S. J., Dubner. 2005. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: William Morrow.
- Lucretius. 1951. *The Nature of the Universe*. Translated by R. E. Latham Harmondsworth, Middlesex, England: Penguin.
- Lurii, A. R. 1976. *Cognitive Development: Its Cultural and Social Foundations*. Cambridge, MA: Harvard University Press.
- Martin, R. M. 1992. *There Are Two Errors in the the Title of This Book: A Sourcebook of Philosophical Puzzles, Problems, and Paradoxes*. Peterborough, Ontario: Broadview Press.
- Nozick, R. 1970. Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel*. N. Rescher, ed., Dordrecht: Reidel.
- Nussbaum, M. C. 1978. *Aristotle's De Motu Animalium: Text with Translation and Notes and Interpretative Essays*. Princeton: Princeton University Press.
- Pascal, B. 1951. *Pensées et Opuscules*. Introd, notices, notes L. Brunschvicg. Paris: Hachette.
- Peng, K., and R. E., Nisbett. 1999. Culture, dialectics, and reasoning about contradiction. *American Psychologist* 54: 741–754.
- Plato. 1997. Meno. In *Complete Works*. Translated by G. Grube J. M. Cooper, ed., 870–96, Indianapolis: Hackett.
- Priest, G. 1997. Sylvan's box: A short story and ten morals. *Notre Dame Journal of Formal Logic* 38(4): 573–582.

- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford, NY: Oxford University Press.
- Ramsey, F. P. 1931. Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. R. B. Braithwaite, ed., preface by G. E. Moore, 52–93, London: Routledge and Kegan Paul.
- Roberts, R. C. 1988. What is an emotion? A sketch. *American Philosophical Quarterly* 97: 183–209.
- Slovic, P., ed. 2000. *The Perception of Risk*. London: Earthscan.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24(2): 1–12.
- Stanovich, K. E. 2004. *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: University of Chicago Press.
- Starr, C. 1969. Societal benefit vs. technological risk. *Science* 165: 1232–1238.
- Strack, F., and R., Deutsch. 2004. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review* 8(3): 220–227.
- Sunstein, C. 2005. *Laws of Fear: Beyond the Precautionary Principle*. New York: Cambridge University Press.
- Sunstein, C. R., and R. H., Thaler. 2003. Libertarian Paternalism Is Not An Oxymoron. *University of Chicago Law Review*, 70(4): 1159–1202.
- Tversky, A., and D., Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–458.
- Whitlow, J. W. J., and W. K., Estes. 1979. Judgment of relative frequency in relation to shifts of event frequency: Evidence for a limited capacity model. *Journal of Experimental Psychology: Human Learning and Memory* 5: 395–408.
- Wilson, T. D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA; London: Harvard University Press, Belnap.
- Zajonc, R. B. 2000. Feeling and thinking: Closing the debate over the independence of affect. In *Feeling and Thinking: The Role of Affect in Social Cognition*. J. P. Forgas, ed., 31–58. Cambridge: Cambridge University Press.

If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide

Paul Slovic

To avoid further disasters, we need political restraint on a world scale. But politics is not the whole story. We have experienced the result of technology in the service of the destructive side of human psychology. Something needs to be done about this fatal combination. The means for expressing cruelty and carrying out mass killing have been fully developed. It is too late to stop the technology. It is to the psychology that we should now turn.

Jonathan Glover, Humanity 2001, p. 144

1 Introduction

“If I look at the mass I will never act. If I look at one, I will.” This statement, uttered by Mother Teresa, captures a powerful and deeply unsettling insight into human nature: Most people are caring and will exert great effort to rescue “the one” whose needy plight comes to their attention. But these same people often become numbly indifferent to the plight of “the one” who is part of a much greater problem. Why does this occur? The answer to this question will help us answer a related question: Why do good people and their governments ignore mass murder and genocide?

There is no simple answer to this question. It is not because we are insensitive to the suffering of our fellow human beings – witness the extraordinary efforts we expend to rescue a person in distress. It is not because we only care about identifiable victims, of similar skin color, who live near us: witness the outpouring of aid to victims of the December 2004 tsunami in South Asia.¹ We cannot simply blame our political leaders. Although President Bush was quite unresponsive to the murder of hundreds of thousands of people in Darfur, it was President Clinton who ignored Rwanda, and President Roosevelt who did little to stop the Holocaust. Behind every

P. Slovic (✉)

Decision Research and Department of Psychology, University of Oregon, Eugene, OR, USA
e-mail: pslovic@darkwinguoregon.edu

¹And, similarly, the aid to Haiti in 2010.

president who ignored mass murder were millions of citizens whose indifference allowed them to get away with it. And it is not only fear of losing American lives in battle that necessarily deters us from acting. We have not even taken quite safe steps that could save many lives, such as bombing the radio stations in Rwanda that were coordinating the slaughter of 800,000 people in 100 days, or supporting the forces of the African Union in Darfur, or just raising our powerful American voices in a threatening shout –*Stop that killing!*—as opposed to turning away in silence.

Every episode of mass murder is distinct and raises unique social, economic, military, and political obstacles to intervention. We therefore recognize that geopolitics, domestic politics, or failures of individual leadership have been important factors in particular episodes. But the repetitiveness of such atrocities, ignored by powerful people and nations, and by the general public, calls for explanations that may reflect some fundamental deficiency in our humanity – a deficiency not in our intentions, but in our very hardware. And a deficiency that, once identified, might possibly be overcome.

One fundamental mechanism that may play a role in many, if not all, episodes of mass-murder neglect involves the capacity to experience *affect*, the positive and negative feelings that combine with reasoned analysis to guide our judgments, decisions, and actions. Research shows that the statistics of mass-murder or genocide, no matter how large the numbers, fail to convey the true meaning of such atrocities. The numbers fail to spark emotion or feeling and thus fail to motivate action. Genocide in Darfur is real, but we do not “feel” that reality. I examine below ways that might make genocide “feel real” and motivate appropriate interventions.

Ultimately, however, I conclude that we cannot only depend on our intuitive feelings about these atrocities but, in addition, we must create and commit ourselves to institutional, legal, and political responses based upon reasoned analysis of our moral obligations to stop the mass annihilation of innocent people.

Although the central focus of this analysis is genocide, the psychological factors underlying affect, imagery, and insensitivity to large-scale harms likely apply as well to damages associated with technology. In particular, the psychological account described here can explain, in part, our failure to respond to the diffuse and seemingly distant threat posed by global warming (see, e.g., Gilbert 2006) as well as the threat posed by the presence of nuclear weaponry.

2 The Lessons of Genocide

Dubinsky (2005, p. 112) reports a news story from *The Gazette* (Montreal; 29 April 1994, at p. A8):

On April 28, 1994: the Associated Press (AP) bureau in Nairobi received a frantic call from a man in Kigali who described horrific scenes of concerted slaughter that had been unfolding in the Rwandan capital ‘every day, everywhere’ for three weeks. ‘I saw people hacked to death, even babies, month-old babies. . . . Anybody who tried to flee was killed in the streets, and people who were hiding were found and massacred.’

Dubinsky (2005, p. 113) further notes that:

The caller’s story was dispatched on the AP newswire for the planet to read, and complemented an OXFAM statement from the same day declaring that the slaughter—the toll of which had already reached 200,000—‘amounts to genocide.’ The following day, U.N. Secretary General Boutros Boutros-Ghali acknowledged the massacres and requested that the Security Council deploy a significant force, a week after the council had reduced the number of U.N. peacekeepers in Rwanda from 2,500 to 270.

Yet the killings continued for another two and a half months. By mid-July, when the government was finally routed by exiled Tutsi rebels, the slaughter had been quelled, and 800,000 were dead, reinforcements from the United Nations were only just arriving.

In his review of the book *Conspiracy to Murder: The Rwandan Genocide* (Melvern 2004), Dubinsky (2005, p. 113) draws an ominous lesson from what happened in Rwanda:

Despite its morally unambiguous heinousness, despite overwhelming evidence of its occurrence (for example, two days into the Rwandan carnage, the U. S. Defense Intelligence Agency possessed satellite photos showing sprawling massacre sites), and despite the relative ease with which it could have been abated (the U.N. commander in Rwanda felt a modest 5,500 reinforcements, had they arrived promptly, could have saved tens of thousands of lives)—despite all this, the world ignored genocide.

Unfortunately, Rwanda is not an isolated incident of indifference to mass murder and genocide. In a deeply disturbing book titled *A Problem from Hell: America and the Age of Genocide*, journalist Samantha Power documents in meticulous detail many of the numerous genocides that occurred during the past century, beginning with the slaughter of two million Armenians by the Turks in 1915 (Power 2003, see Table 1). In every instance, American response was inadequate. She concludes, “No U. S. president has ever made genocide prevention a priority, and no U. S. president has ever suffered politically for his indifference to its occurrence. It is thus no coincidence that genocide rages on” (Power 2003, p. xxi).

A second lesson to emerge from the study of genocide is that media news coverage is similarly inadequate. The past century has witnessed a remarkable transformation in the ability of the news media to learn about, and report on, world events. The vivid, dramatic coverage of the December 2004 Tsunami in South Asia and the similarly intimate and exhaustive reporting of the destruction of lives and property by Hurricane Katrina in September 2005 demonstrate how thorough and

Table 1 A century of genocide

Armenia (1915)
Ukraine (1932–1933)
Nazi Germany/Holocaust (World War II)
Bangladesh (1971)
Cambodia (1975–1979)
Countries in the former Yugoslavia (1990s)
Rwanda (1994)
Zimbabwe (2000)
Congo (Today)
Darfur (Today)
? (Tomorrow)

how powerful news coverage of humanitarian disasters can be. But the intense coverage of recent natural disasters stands in sharp contrast to the lack of reporting on the ongoing genocides in Darfur and other regions in Africa, in which hundreds of thousands of people have been murdered and millions forced to flee their burning villages and relocate in refugee camps. According to the Tyndall Report, which monitors U. S. television coverage, ABC news allotted a total of 18 minutes on the Darfur genocide in its nightly newscasts in 2004, NBC had only five minutes, and CBS only three minutes. Martha Stewart and Michael Jackson received vastly greater coverage, as did Natalee Holloway, the American girl missing in Aruba. With the exception of the relentless reporting by *New York Times* columnist Nicholas Kristof, the print media have done little better in covering Darfur.

Despite lack of attention by the news media, U. S. government officials have known of the mass murders and genocides that took place during the past century. Power (2003) attempts to explain the failure to act on that knowledge as follows:

... the atrocities that were known remained abstract and remote... Because the savagery of genocide so defies our everyday experience, many of us failed to *wrap our minds around it*... Bystanders were thus able to retreat to the ‘twilight between knowing and not knowing.’ (p. 505, italics added)

I shall argue below that the disengagement exemplified by failing to “wrap our minds” around genocide and retreating to the “twilight between knowing and not knowing” is at the heart of our failure to act against genocide. Samantha Power’s insightful explanation is supported by the research literature in cognitive and social psychology, as described in the sections to follow.

3 Lessons from Psychological Research

In 1994, Roméo Dallaire, the commander of the tiny U.N. peacekeeping mission in Rwanda, was forced to watch helplessly as the slaughter he had foreseen and warned about began to unfold. Writing of this massive humanitarian disaster a decade later he encouraged scholars “to study this human tragedy and to contribute to our growing understanding of the genocide. If we do not understand what happened, how will we ever ensure it does not happen again?” Dallaire (2005, p. 548).

Researchers in psychology, economics, and a multidisciplinary field called behavioral decision theory have developed theories and findings that, in part, begin to explain the pervasive neglect of genocide.

3.1 Affect, Attention, Information, and Meaning

My search to identify a fundamental deficiency in human psychology that causes us to ignore mass murder and genocide has led to a theoretical framework that describes the importance of emotions and feelings in guiding decision making and

behavior. Perhaps the most basic form of feeling is affect, the sense (not necessarily conscious) that something is good or bad. Affective responses occur rapidly and automatically – note how quickly you sense the feelings associated with the word “treasure” or the word “hate.” A large research literature in psychology documents the importance of affect in conveying meaning upon information and motivating behavior (Barrett and Salovey 2002; Clark and Fiske 1982; Forgas 2000; Ledoux 1996; Mowrer 1960; Tomkins 1962, 1963; Zajonc 1980). Without affect, information lacks meaning and won’t be used in judgment and decision making (Loewenstein et al. 2001; Slovic et al. 2002).

Affect plays a central role in what have come to be known as “dual-process theories” of thinking. As Seymour Epstein (1994) has observed: “There is no dearth of evidence in every day life that people apprehend reality in two fundamentally different ways, one variously labeled intuitive, automatic, natural, non-verbal, narrative, and experiential, and the other analytical, deliberative, verbal, and rational” (p. 710).

Table 2, adapted from Epstein, further compares these two systems, which Stanovich and West (2000) labeled *System 1* and *System 2*. One of the characteristics of the experiential system is its affective basis. Although analysis is certainly important in many decision-making circumstances, reliance on affect and emotion is generally a quicker, easier, and more efficient way to navigate in a complex, uncertain and sometimes dangerous world. Many theorists have given affect a direct and primary role in motivating behavior. Epstein’s (1994) view on this is as follows:

The experiential system is assumed to be intimately associated with the experience of affect... which refer[s] to subtle feelings of which people are often unaware. When a person responds to an emotionally significant event... The experiential system automatically searches its memory banks for related events, including their emotional accompaniments... If the activated feelings are pleasant, they motivate actions and thoughts anticipated to reproduce the feelings. If the feelings are unpleasant, they motivate actions and thoughts anticipated to avoid the feelings. (p. 716)

Table 2 Two modes of thinking: comparison of experiential and analytic systems

System 1: Experiential system	System 2: Analytic system
Affective: pleasure-pain oriented	Logical: reason oriented (what is sensible)
Connections by association	Connections by logical assessment
Behavior mediated by feelings from past experiences	Behavior mediated by conscious appraisal of events
Encodes reality in images, metaphors, and narratives	Encodes reality in abstract symbols, words, and numbers
More rapid processing: oriented toward immediate action	Slower processing: oriented toward delayed action
Self-evidently valid: “experiencing is believing”	Requires justification via logic and evidence

Source: Adapted from Epstein (1994).

Underlying the role of affect in the experiential system is the importance of images, to which positive or negative feelings become attached. Images in this system include not only visual images, important as these may be, but words, sounds, smells, memories, and products of our imagination.

In his Nobel Prize Address, Daniel Kahneman notes that the operating characteristics of System 1 are similar to those of human perceptual processes (Kahneman 2003). He points out that one of the functions of System 2 is to monitor the quality of the intuitive impressions formed by System 1. Kahneman and Frederick (2002) suggest that this monitoring is typically rather lax and allows many intuitive judgments to be expressed in behavior, including some that are erroneous. This point has important implications that will be discussed later.

In addition to positive and negative affect, more nuanced feelings such as empathy, sympathy, compassion, sadness, pity, and distress have been found to be critical for motivating people to help others (Coke et al. 1978; Eisenberg and Miller 1987). As Batson (1990, p. 339) put it, “. . . considerable research suggests that we are more likely to help someone in need when we ‘feel for’ that person. . .”

One last important psychological element in this story is attention. Just as feelings are necessary for motivating helping, attention is necessary for feelings. Research shows that attention magnifies emotional responses to stimuli that are already emotionally charged (Fenske and Raymond 2006; Vuilleumier et al. 2003). The psychological story can be summarized by the diagram in Fig. 1. Research to be described in this paper demonstrates that imagery and feeling are lacking when large losses of life are represented simply as numbers or statistics. Other research shows that attention is greater for individuals and loses focus and intensity when targeted at groups of people (Hamilton and Sherman 1996; Susskind et al. 1999). The foibles of imagery and attention impact feelings in a manner that can help explain apathy toward genocide.

Although the model sketched in Fig. 1 could incorporate elements of System 1 thinking, System 2 thinking, or both, a careful analysis by Haidt (2001) gives priority to System 1. Haidt argues that moral intuitions (akin to System 1) precede moral judgments. Specifically, he asserts that

“. . . moral intuition can be defined as the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike) without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion. Moral intuition is therefore. . . akin to aesthetic judgment. One sees or hears about a social event and one instantly feels approval or disapproval” (p. 818; see also Hume 1777/1960 for an earlier version of this argument).

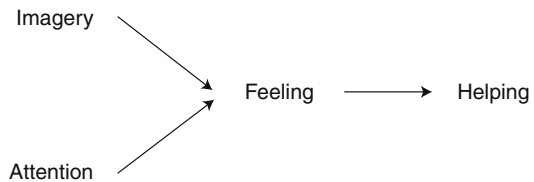
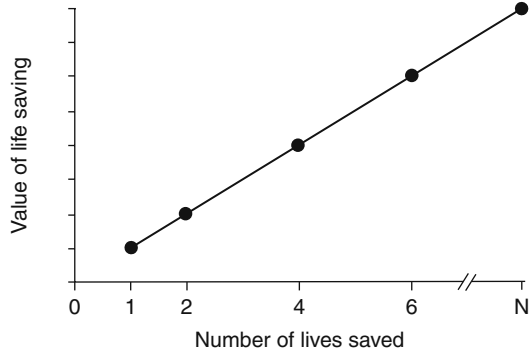


Fig. 1 Imagery and attention produce feelings that motivate helping behavior

4 Affect, Analysis, and the Value of Human Lives

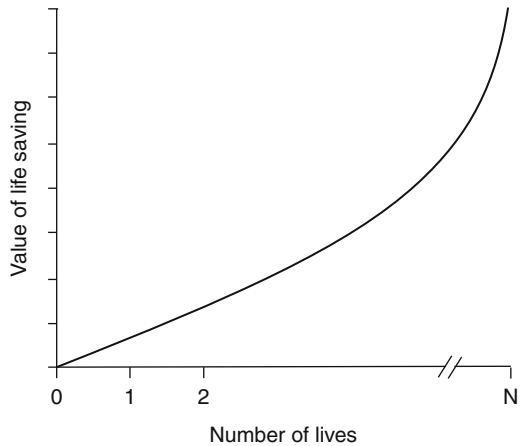
How *should* we value the saving of human lives? If we believe that every human life is of equal value (a view likely endorsed by System 2 thinking), the value of saving N lives is N times the value of saving one life, as represented by the linear function in Fig. 2.

Fig. 2 A normative model for valuing the saving of human lives. Every human life is of equal value



An argument can also be made for a model in which large losses of life are disproportionately more serious because they threaten the social fabric and viability of a community as depicted in Fig. 3.

Fig. 3 Another normative model: Large losses threaten the viability of the group or society (as with genocide)



How *do* we actually value humans lives? I shall present evidence in support of two descriptive models linked to affect and System 1 thinking that reflect values for lifesaving profoundly different from the normative models shown in Figs. 1 and 2. Both of these models are instructive with regard to apathy toward genocide.

4.1 The Psychophysical Model

Affect is a remarkable mechanism that enabled humans to survive the long course of evolution. Before there were sophisticated analytic tools such as probability theory, scientific risk assessment, and cost/benefit calculus, humans used their senses, honed by experience, to determine whether the animal lurking in the bushes was safe to approach or the murky water in the pond was safe to drink. Simply put, System 1 thinking evolved to protect individuals and their small family and community groups from present, visible, immediate dangers. This affective system did not evolve to help us respond to distant, mass murder. As a result, System 1 thinking responds to large-scale atrocities in ways that are less than desirable.

Fundamental qualities of human behavior are, of course, recognized by others besides scientists. American writer Annie Dillard cleverly demonstrates the limitation of our affective system as she seeks to help us understand the humanity of the Chinese nation: “There are 1,198,500,000 people alive now in China. To get a *feel* for what this *means*, simply take yourself – in all your singularity, importance, complexity, and love – and multiply by 1,198,500,000. See? Nothing to it” (Dillard 1999, p. 47, italics added).

We quickly recognize that Dillard is joking when she asserts “nothing to it.” We know, as she does, that we are incapable of *feeling* the humanity behind the number 1,198,500,000. The circuitry in our brain is not up to this task. This same incapacity is echoed by Nobel prize winning biochemist Albert Szent Gyorgi as he struggles to comprehend the possible consequences of nuclear war: “I am deeply moved if I see one man suffering and would risk my life for him. Then I talk impersonally about the possible pulverization of our big cities, with a hundred million dead. I am unable to multiply one man’s suffering by a hundred million.”

There is considerable evidence that our affective responses and the resulting value we place on saving human lives may follow the same sort of “psychophysical function” that characterizes our diminished sensitivity to a wide range of perceptual and cognitive entities – brightness, loudness, heaviness, and money – as their underlying magnitudes increase.

What psychological principles lie behind this insensitivity? In the nineteenth century, E. H. Weber and Gustav Fechner discovered a fundamental psychophysical principle that describes how we perceive changes in our environment. They found that people’s ability to detect changes in a physical stimulus rapidly decreases as the magnitude of the stimulus increases (Weber 1834; Fechner 1860). What is known today as “Weber’s law” states that in order for a change in a stimulus to become *just noticeable*, a fixed percentage must be added. Thus, perceived difference is a relative matter. To a small stimulus, only a small amount must be added to be noticeable. To a large stimulus, a large amount must be added. Fechner proposed a logarithmic law to model this nonlinear growth of sensation. Numerous empirical studies by S. S. Stevens (1975) have demonstrated that the growth of sensory magnitude (ψ) is best fit by a power function of the stimulus magnitude Φ ,

$$\psi = k\Phi^\beta,$$

where the exponent β is typically less than one for measurements of phenomena such as loudness, brightness, and even the value of money (Galanter 1962). For example, if the exponent is 0.5 as it is in some studies of perceived brightness, a light that is four times the intensity of another light will be judged only twice as bright.

Our cognitive and perceptual systems seem to be designed to sensitize us to small changes in our environment, possibly at the expense of making us less able to detect and respond to large changes. As the psychophysical research indicates, constant increases in the magnitude of a stimulus typically evoke smaller and smaller changes in response. Applying this principle to the valuing of human life suggests that a form of *psychophysical numbing* may result from our inability to appreciate losses of life as they become larger (see Fig. 4). The function in Fig. 4 represents a value structure in which the importance of saving one life is great when it is the first, or only, life saved, but diminishes marginally as the total number of lives saved increases. Thus, psychologically, the importance of saving one life is diminished against the background of a larger threat – we will likely not “feel” much different, nor value the difference, between saving 87 lives and saving 88, if these prospects are presented to us separately.

Kahneman and Tversky (1979) have incorporated this psychophysical principle of decreasing sensitivity into prospect theory, a descriptive account of decision making under uncertainty. A major element of prospect theory is the value function, which relates subjective value to actual gains or losses. When applied to human lives, the value function implies that the subjective value of saving a specific number of lives is greater for a smaller tragedy than for a larger one.

Fetherstonhaugh et al. (1997) documented this potential for diminished sensitivity to the value of life – i.e., “psychophysical numbing” – by evaluating people’s willingness to fund various lifesaving medical treatments. In a study involving a hypothetical grant funding agency, respondents were asked to indicate the number of lives a medical research institute would have to save to merit receipt of a \$10 million grant. Nearly two-thirds of the respondents raised their minimum benefit requirements to warrant funding when there was a larger at-risk population, with a median value of 9,000 lives needing to be saved when 15,000 were at risk,

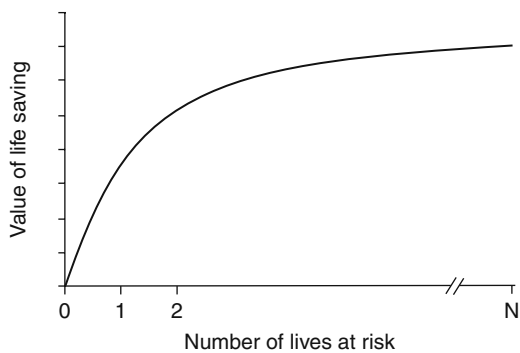


Fig. 4 A psychophysical model describing how the saving of human lives may actually be valued

compared to a median of 100,000 lives needing to be saved out of 290,000 at risk. By implication, respondents saw saving 9,000 lives in the “smaller” population as more valuable than saving ten times as many lives in the largest.

Several other studies in the domain of life-saving interventions have documented similar psychophysical numbing or proportional reasoning effects (Baron 1997; Bartels and Burnett 2006; Fetherstonhaugh et al. 1997; Friedrich et al. 1999; Jenni and Loewenstein 1997; Ubel et al. 2001). For example, Fetherstonhaugh et al. (1997) also found that people were less willing to send aid that would save 1500 lives in Rwandan refugee camps as the size of the camps’ at-risk population increased. Friedrich et al. (1999) found that people required more lives to be saved to justify mandatory antilock brakes on new cars when the alleged size of the at-risk pool (annual braking-related deaths) increased.

These diverse strategies of lifesaving demonstrate that the *proportion* of lives saved often carries more weight than the *number* of lives saved when people evaluate interventions. Thus, extrapolating from Fetherstonhaugh et al., one would expect that, in separate evaluations, there would be more support for saving 80% of 100 lives at risk than for saving 20% of 1,000 lives at risk. This is consistent with an affective (System 1) account, in which the number of lives saved conveys little affect but the proportion saved carries much feeling: 80% is clearly “good” and 20% is “poor.”

Slovic et al. (2004), drawing upon the finding that proportions appear to convey more feeling than do numbers of lives, predicted (and found) that college students, in a between-groups design, would more strongly support an airport-safety measure expected to save 98% of 150 lives at risk than a measure expected to save 150 lives. Saving 150 lives is diffusely good, and therefore somewhat hard to evaluate, whereas saving 98% of something is clearly very good because it is so close to the upper bound on the percentage scale, and hence is highly weighted in the support judgment. Subsequent reduction of the percentage of 150 lives that would be saved to 95, 90, and 85% led to reduced support for the safety measure but each of these

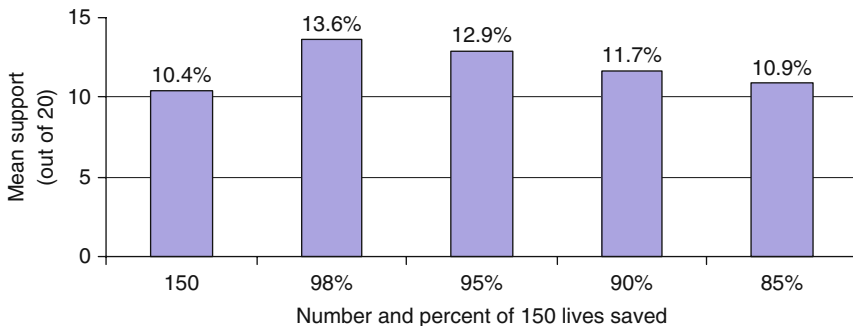


Fig. 5 Airport safety study: Saving a percentage of 150 lives receives higher support ratings than does saving 150 lives. *Note.* Bars describe mean responses to the question, “How much would you support the proposed measure to purchase the new equipment?” The response scale ranged from 0 (*would not support at all*) to 20 (*very strong support*; Slovic et al. 2002)

percentage conditions still garnered a higher mean level of support than did the Save 150 Lives Condition (Fig. 5).

This research on psychophysical numbing is important because it demonstrates that feelings necessary for motivating lifesaving actions are not congruent with the normative models in Figs. 2 and 3. The nonlinearity displayed in Fig. 4 is consistent with the disregard of incremental loss of life against a background of a large tragedy. However it does not fully explain the utter collapse of compassion represented by apathy toward genocide because it implies that the response to initial loss of life will be strong and maintained as the losses increase. Evidence for a second descriptive model, one better suited to explain the collapse of compassion, follows.

5 Numbers and Numbness: Images and Feeling

The behavioral theories and data confirm what keen observers of human behavior have long known. Numerical representations of human lives do not necessarily convey the importance of those lives. All too often the numbers represent dry statistics, “human beings with the tears dried off,” that lack feeling and fail to motivate action (Slovic and Slovic 2004). How can we impart the feelings that are needed for rational action? There have been a variety of attempts to do this that may be instructive. Most of these involve highlighting the images that lie beneath the numbers. As nature writer and conservationist Rick Bass (1996) observes in his plea to conserve the Yaak Valley in Montana,

The numbers are important, and yet they are not everything. For whatever reasons, images often strike us more powerfully, more deeply than numbers. We seem unable to hold the emotions aroused by numbers for nearly as long as those of images. We quickly grow numb to the facts and the math. (p. 87)

Images seem to be the key to conveying affect and meaning, though some imagery is more powerful than others. After struggling to appreciate the mass of humanity in China, Annie Dillard turned her thoughts to April 30, 1991, when 138,000 people drowned in Bangladesh. At dinner, she mentions to her daughter – 7 years old – that it is hard to imagine 138,000 people drowning. “No, it’s easy,” says her daughter. “Lots and lots of dots in blue water” (Dillard 1999, p. 131). Again we are confronted with impoverished meaning associated with large losses of life.

Other images may be more effective. Organizers of a rally designed to get Congress to do something about 38,000 deaths a year from handguns piled 38,000 pairs of shoes in a mound in front of the Capitol (Associated Press 1994). Students at a middle school in Tennessee, struggling to comprehend the magnitude of the holocaust, collected 6 million paper clips as a centerpiece for a memorial (Schroeder and Schroeder-Hildebrand 2004).

Probably the most important image to represent a human life is that of a single human face. Journalist Paul Neville writes about the need to probe beneath the statistics of joblessness, homelessness, mental illness, and poverty in his home state of Oregon, in order to discover the people behind the numbers – who they are, what

they look like, how they sound, what they feel, what hopes and fears they harbor. He concludes: “I don’t know when we became a nation of statistics. But I know that the path to becoming a nation – and a community – of people, is remembering the faces behind the numbers” (Neville 2004). After September 11, 2001, many newspapers published biographical sketches of the victims, with photos, a dozen or so each day until all had been featured.

When it comes to eliciting compassion, the identified individual victim, with a face and a name, has no peer. Psychological experiments demonstrate this clearly but we all know it as well from personal experience and media coverage of heroic efforts to save individual lives. One of the most publicized events occurred when an 18-month-old child, Jessica McClure, fell 22 feet into a narrow abandoned well shaft. The world watched tensely as rescuers worked for 2½ days to rescue her. Almost two decades later, the joyous moment of Jessica’s rescue is portrayed with resurrection-like overtones on a website devoted to pictures of the event (see Fig. 6).

But the face need not even be human to motivate powerful intervention. In 2001, an epidemic of foot and mouth disease raged throughout the United Kingdom.



Fig. 6 The rescue of baby Jessica. Source: “The Baby Jessica Rescue Web Page,” <http://www.caver.net/j/jrescue.html>. Accessed 24 November 2008

Millions of cattle were slaughtered to stop the spread. The disease waned and animal rights activists demanded an end to further killing. But the killings continued until a newspaper photo of a cute 12-day-old calf named Phoenix being targeted for slaughter led the government to change its policy. Individual canine lives are highly valued, too. A dog stranded aboard a tanker adrift in the Pacific was the subject of one of the most costly animal rescue efforts ever. An Associated Press article discloses that the cost of rescue attempts had already reached \$48,000 and the Coast Guard was prepared to spend more, while critics charged that the money could be better spent on children that go to bed hungry (Song 2002).

In a bizarre incident that, nonetheless, demonstrates the special value of an individual life, an article in the BBC News online edition of November 19, 2005, reports the emotional response in the Netherlands to the shooting of a sparrow that trespassed onto the site of a domino competition and knocked over 23,000 tiles. A tribute website was set up and attracted tens of thousands of hits. The head of the Dutch Bird Protection Agency, appearing on television, said that though it was a very sad incident, it had been blown out of all proportion. "I just wish we could channel all this energy that went into one dead sparrow into saving the species," he said (BBC News 2005).

Going beyond faces, names, and other simple images, writers and artists have long recognized the power of narrative to bring feelings and meaning to tragedy. Barbara Kingsolver (1995) makes this point eloquently in her book *High Tide in Tucson*.

The power of fiction is to create empathy. It lifts you away from your chair and stuffs you gently down inside someone else's point of view. . . . A newspaper could tell you that one hundred people, say, in an airplane, or in Israel, or in Iraq, have died today. And you can think to yourself, "How very sad," then turn the page and see how the Wildcats fared. But a novel could take just one of those hundred lives and show you exactly how it felt to be that person rising from bed in the morning, watching the desert light on the tile of her doorway and on the curve of her daughter's cheek. You could taste that person's breakfast, and love her family, and sort through her worries as your own, and know that a death in that household will be the end of the only life that someone will ever have. As important as yours. As important as mine. (p. 231)

Showing insight into the workings of our affective system as keen as any derived from the psychologist's laboratory, Kingsolver continues:

Confronted with knowledge of dozens of apparently random disasters each day, what can a human heart do but slam its doors? No mortal can grieve that much. We didn't evolve to cope with tragedy on a global scale. Our defense is to pretend there's no thread of event that connects us, and that those lives are somehow not precious and real like our own. It's a practical strategy, to some ends, but the loss of empathy is also the loss of humanity, and that's no small tradeoff.

Art is the antidote that can call us back from the edge of numbness, restoring the ability to feel for another. (p. 231–232)

Although Kingsolver is describing the power of fiction, nonfiction narrative can be just as effective. *The Diary of Anne Frank* and Elie Wiesel's *Night* certainly convey, in a powerful way, the meaning of the Holocaust statistic "six million dead."

6 The Collapse of Compassion

Vivid images of recent natural disasters in South Asia and the American Gulf Coast, and stories of individual victims, brought to us through relentless, courageous, and intimate news coverage, certainly unleashed a tidal wave of compassion and humanitarian aid from all over the world. Private donations to the victims of the December 2004 tsunami exceeded \$1 billion. Charities such as Save the Children have long recognized that it is better to endow a donor with a single, named child to support than to ask for contributions to the bigger cause. Perhaps there is hope that vivid, personalized media coverage of genocide could motivate intervention.

Perhaps. But again we should look to research to assess these possibilities. Numerous experiments have demonstrated the “ identifiable victim effect” which is also so evident outside the laboratory. People are much more willing to aid identified individuals than unidentified or statistical victims (Kogut and Ritov 2005a; Schelling 1968; Small and Loewenstein 2003, 2005; Jenni and Loewenstein 1997). Small et al. (2007) gave people leaving a psychological experiment the opportunity to contribute up to \$5 of their earnings to Save the Children. The study consisted of three separate conditions: (1) identifiable victim, (2) statistical victims, and (3) identifiable victim with statistical information. The information provided for the identifiable and statistical conditions is shown in Fig. 7. Participants in each condition were told that “any money donated will go toward relieving the severe food crisis in Southern Africa and Ethiopia.” The donations in fact went to Save the Children, but they were earmarked specifically for Rokia in Conditions 1 and 3 and not specifically earmarked in Condition 2. The average donations are presented in Fig. 8. Donations in response to the identified individual, Rokia, were


<i>Statistical Lives</i>	<ul style="list-style-type: none"> ● Food shortages in Malawi are affecting more than 3 million children. ● In Zambia, severe rainfall deficits have resulted in a 42 percent drop in maize production from 2000. As a result, an estimated 3 million Zambians face hunger. ● Four million Angolans — one third of the population — have been forced to flee their homes. ● More than 11 million people in Ethiopia need immediate food assistance.
<i>Identifiable Lives</i>	<p>Any money that you donate will go to Rokia, a 7-year-old girl from Mali, Africa. Rokia is desperately poor, and faces a threat of severe hunger or even starvation. Her life will be changed for the better as a result of your financial gift. With your support, and the support of other caring sponsors, Save the Children will work with Rokia's family and other members of the community to help feed her, provide her with education, as well as basic medical care and hygiene education.</p>
	

Fig. 7 Donating money to save statistical and identified lives. Reprinted from Small et al. (2007). Copyright (2006), with permission from Elsevier

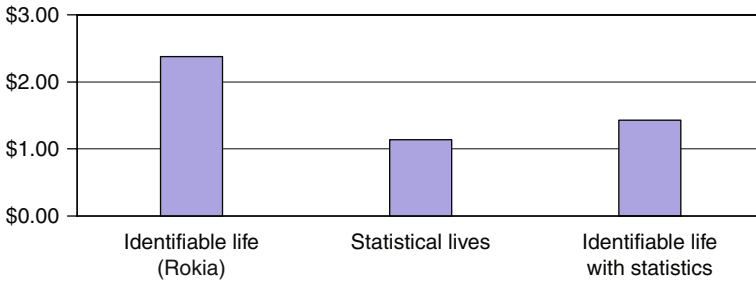


Fig. 8 Mean donations. Reprinted from Small et al. (2007), Copyright (2006), with permission from Elsevier

far greater than donations in response to the statistical portrayal of the food crisis. Most important, however, and most discouraging, was the fact that coupling the statistical realities with Rokia’s story significantly *reduced* the contributions to Rokia. Alternatively, one could say that using Rokia’s story to “put a face behind the statistical problem” did not do much to increase donations (the difference between the mean donations of \$1.43 and \$1.14 was not statistically reliable).

Small et al. also measured feelings of sympathy toward the cause (Rokia or the statistical victims). These feelings were most strongly correlated with donations when people faced an identifiable victim.

A follow-up experiment by Small et al. provided additional evidence for the importance of feelings. Before being given an opportunity to donate, study participants were either primed to feel (“Describe your feelings when you hear the word ‘baby,’” and similar items) or to answer five questions such as “If an object travels at five feet per minute, then by your calculations how many feet will it travel in 360 seconds?” Priming analytic thinking (calculation) reduced donations to the identifiable victim (Rokia) relative to the feeling-based thinking prime. Yet the two primes had no distinct effect on statistical victims, which is symptomatic of the difficulty in generating feelings for such victims.

Annie Dillard reads in her newspaper the headline “Head Spinning Numbers Cause Mind to Go Slack.” She struggles to think straight about the great losses that the world ignores: “More than two million children die a year from diarrhea and eight hundred thousand from measles. Do we blink? Stalin starved seven million Ukrainians in 1 year, Pol Pot killed two million Cambodians. . .” She writes of “compassion fatigue” and asks, “At what number do other individuals blur for me?” (Dillard 1999, pp. 130–131).

An answer to Dillard’s question is beginning to emerge from behavioral research. Studies by Hamilton and Sherman (1996); Susskind et al. (1999) find that a single individual, unlike a group, is viewed as a psychologically coherent unit. This leads to more extensive processing of information and clearer impressions about individuals than about groups. Kogut and Ritov (2005b) hypothesized that the processing of information related to a single victim might be fundamentally different from the processing of information concerning a group of victims. They predicted that people

will tend to feel more distress and compassion when considering an identified single victim than when considering a group of victims, even if identified, resulting in a greater willingness to help the identified individual victim.

Kogut and Ritov (2005a, 2005b) tested their predictions in a series of studies in which participants were asked to contribute to a costly life-saving treatment needed by a sick child or a group of eight sick children. The target amount needed to save the child (children) was the same in both conditions, 1.5 million Israeli Shekels (about \$300,000). All contributions were actually given to an organization that helps children with cancer. In addition to deciding whether or how much they wanted to contribute, participants in some studies rated their feelings of distress (feeling worried, upset, and sad) towards the sick child (children).

The mean contributions to the group of eight and to the individuals taken from the group are shown in Fig. 9 for one of the studies by Kogut and Ritov (2005b). Contributions to the individuals in the group, as individuals, were far greater than were contributions to the entire group. In a separate study, ratings of distress (not shown in the figure) were also higher in the individual condition.

But could the results in Fig. 9 be explained by the possibility that donors believed that families in the group condition would have an easier time obtaining the needed money which, in fact, was less per child in that condition? Further testing ruled out this explanation. For example, Kogut and Ritov asked people to choose between donating to a single child of the eight or donating to the remaining seven children. Many more (69%) chose to donate to the group, demonstrating a sensitivity to the number of victims in need that was not evident in the noncomparative evaluations. Kogut and Ritov concluded that the greater donations to the single victim most likely stem from the stronger emotions evoked by such victims in conditions where donors evaluated only a single child or only the group.

Recall Samantha Power's assertion that those who know about genocide somehow "fail to wrap their minds around it." Perhaps this is a layperson's terminology for the less coherent processing of information about groups observed by Hamilton and Sherman (1996) and Susskind et al. (1999). And perhaps the beginning of this failure is evident with as few as eight victims.

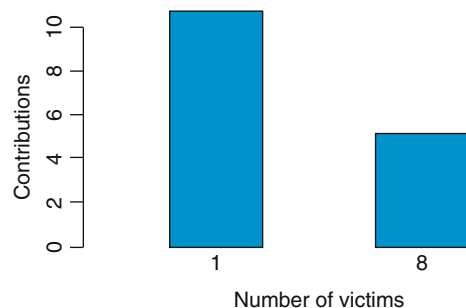


Fig. 9 Mean contributions to individuals and their group. Reprinted from Kogut and Ritov (2005b), Copyright (2005), with permission from Elsevier

Or, perhaps the deterioration of compassion may appear in groups as small as two persons! A recent study suggests this. Västfjäll et al. (2009) decided to test whether the effect found by Kogut and Ritov would occur as well for donations to two starving children. Following the protocol designed by Small et al. (2007), they gave one group of Swedish students the opportunity to contribute their earnings from another experiment to Save the Children to aid Rokia, whose plight was described as in Fig. 7. A second group was offered the opportunity to contribute their earnings to Save the Children to aid Moussa, a seven-year-old boy from Mali (photograph provided) who was similarly described as in need of food aid. A third group was shown the vignettes and photos of Rokia and Moussa and was told that any donation would go to both of them, Rokia *and* Moussa. The donations were real and were sent to Save the Children. Participants also rated their feelings about donating on a 1 (*negative*) to 5 (*positive*) scale. Affect was found to be least positive in the combined condition and donations were smaller in that condition (see Fig. 10). In the individual-child conditions, the size of the donation made was strongly correlated with rated feelings ($r = 0.52$ for Rokia; $r = 0.52$ for Moussa). However this correlation was much reduced ($r = 0.19$) in the combined condition.

As unsettling as is the valuation of life-saving portrayed by the psychophysical model in Fig. 4, the studies just described suggest an even more disturbing psychological tendency. Our capacity to feel is limited. To the extent that valuation of life-saving depends on feelings driven by attention or imagery (recall Fig. 1), it might follow the function shown in Fig. 11, where the emotion or affective feeling is greatest at $N = 1$ but begins to decline at $N = 2$ and collapses at some higher value of N that becomes simply “a statistic.” In other words, returning to Annie Dillard’s worry about compassion fatigue, perhaps the “blurring” of individuals begins at two! Whereas Lifton (1967) coined the term “psychic numbing” to describe the “turning off” of feeling that enabled rescue workers to function during the horrific aftermath of the Hiroshima bombing, Fig. 11 depicts a form of numbing that is not beneficial.

Feelings and donations decline at $N = 2$!

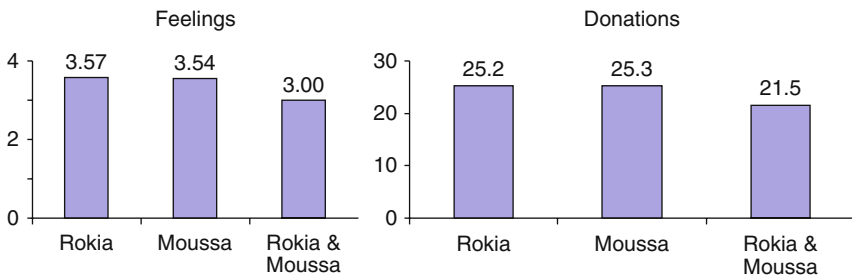


Fig. 10 Mean affect ratings (*left*) and mean donations (*right*) for individuals and their combination. Source: Västfjäll et al. (2009)

Fig. 11 A model depicting psychic numbing—the collapse of compassion—when valuing the saving of lives



Rather, it leads to apathy and inaction, consistent with what is seen repeatedly in response to mass murder and genocide.

7 The Mournful Math of Darfur: The Dead Don't Add Up

The title of this section comes from the headline in a *New York Times* article (Lacey 2005) describing the difficulty that officials are having in determining the actual death toll in Darfur. The diverse and savage methods of killing defy accurate accounting, with estimates at the time of the article ranging between 60,000 and 400,000. The point I have been arguing in this paper, that the numbers don't really matter because we are insensitive to them, is obviously not appreciated by those struggling to tally the dead. They are described as "... engaging in guesswork for a cause. They say they are trying to count the deaths to shock the world into stopping the number from rising higher..." An American professor leading the accounting effort on behalf of the Coalition for International Justice argues that calculating the death toll is important to "... focus the attention of people... to give them some sense of the scale of what's happening in Darfur."

If those attempting to count the dead are naïve about the impact the numbers may have, the writer of the story is not. He concludes:

... eventually, when Darfur's violence mercifully ends, a number will be agreed upon. That number, like the figure of 800,000 for the Rwanda massacre, will be forever appended to the awful events. The rest of the world, slow to react to Darfur, will then have plenty of opportunity to think about it, and wonder why it was able to grow as large as it did. (Lacey 2005)

8 Facing Genocide

Clearly there are political obstacles posing challenges to those who would consider intervention in genocide, and physical risks as well. What I have tried to describe in this paper are the formidable psychological obstacles centered around the difficulties

in wrapping our minds around genocide and forming the emotional connections to its victims that are necessary to motivate us to overcome these other obstacles.

Are we destined to stand numbly and do nothing as genocide rages on for another century? Can we overcome the psychological obstacles to action? There are no simple solutions. One possibility is to infuse System 1 with powerful affective imagery such as that associated with Katrina and the South Asian tsunami. This would require pressure on the media to do its job and report the slaughter of thousands of innocent people aggressively and vividly, as though it were real news. Nicholas Kristof, a columnist for the *New York Times*, has provided a model to emulate for his persistent and personalized reporting of the genocide in Darfur, but he is almost a lone voice in the mainstream U. S. media. Another way to engage our experiential system would be to bring people from Darfur into our communities and our homes to tell their stories.

But, as powerful as System 1 is, when infused with vivid experiential stimulation (witness the moral outrage triggered by the photos of abuse at the Abu Ghraib prison in Iraq), it has a darker side. We cannot rely on it. It depends upon attention and feelings that may be hard to arouse and sustain over time for large numbers of victims, not to speak of numbers as small as two. Left to its own devices, System 1 will likely favor individual victims and sensational stories that are closer to home and easier to imagine. It will be distracted by images that produce strong, though erroneous, feelings, like percentages as opposed to actual numbers. Our sizable capacity to care for others may also be overridden by more pressing personal interests. Compassion for others has been characterized by Batson et al. (1983) as “a fragile flower, easily crushed by self-concern” (p. 718). Faced with genocide, we cannot rely on our moral intuitions alone to guide us to act properly.

A more promising path might be to force System 2 to play a stronger role, not just to provide us with reasons why genocide is wrong – these reasons are obvious and System 1 will appropriately sense their moral messages (Haidt 2001). As Kahneman (2003) argues, one of the important functions of System 2 is to monitor the quality of mental operations and overt behaviors produced by System 1 (see also Gilbert 2002; Stanovich and West 2002).

Most directly, deliberate analysis of the sobering messages contained in this paper should make it clear that we need to create laws and institutions that will *compel* appropriate action when information about genocide becomes known. However, such precommitted response is not as easy as it might seem. Shortly after World War II, on December 9, 1948, the U. N. General Assembly drafted and adopted the Convention for the Prevention and Punishment of the Crime of Genocide. Hopes were high as the world’s states committed themselves to “liberate mankind from such an odious scourge” as genocide (Convention preamble). Yet it took 40 years for the United States to ratify a watered-down version of this treaty, which has been honored mostly in its breach (Power 2003; Schabas 1999). Objections have centered around lack of clarity in the definition of genocide, including the numerical criteria necessary to trigger action. Some feared that the act would be used to target Americans unjustly. Senator William Proxmire took up the cause in 1967, making 3,211 speeches in support of ratification over a 19-year period. However, only

Ronald Reagan's backing, to atone for his politically embarrassing visit to a cemetery in Germany where officials of the Nazi SS were buried, tipped the political balance toward ratification in 1988 of a weakened version of the Convention. When the United States had its first chance to use the law to stop the destruction of Iraq's rural Kurdish population, special interests, economic profit, and political concerns led the Reagan administration to side instead with the genocidal regime of Saddam Hussein (Power 2003).

In this paper I have drawn upon common observation and behavioral research to argue that we cannot depend only upon our moral feelings to motivate us to take proper actions against genocide. That places the burden of response squarely upon the shoulders of moral argument and international law. The genocide convention was supposed to meet this need, but it has not been effective. It is time to reexamine this failure in light of the psychological deficiencies described here and design legal and institutional mechanisms that will enforce proper response to genocide and other crimes against humanity.²

9 Postscript

Roméo Dallaire, in recounting the anguishing story of his failure to convince the United Nations to give him the mandate and force to stop the impending slaughter in Rwanda observes that, "... at its heart, the Rwandan story is the story of the failure of humanity to heed the call for help from an endangered people" (Dallaire 2005, p. 516).

The political causes of this and other such failures are rather well known. What I have tried to describe here are the psychological factors that allow politics to trump morality.

Dallaire (2005) challenges his readers with several questions: "Are we all human, or are some more human than others? If we believe that all humans are human, then how are we going to prove it? It can only be proven through our actions" (p. 522).

A final image: President George W. Bush stands by the casket of Rosa Parks in the rotunda of the U. S. Capitol, paying his respects. Why did the President and the nation so honor this woman? Because, by refusing to give up her seat on the bus she courageously asserted her humanity, answering Dallaire's questions by her actions. At almost the same time as the nation was honoring Parks, the U. S. Congress was stripping \$50 million from the Foreign Operations Bill that was to help pay for

²A thoughtful reviewer of this paper questions my focus on preventing genocide. The reviewer asserts that numbers of preventable deaths from poverty, starvation, and disease are far larger than the numbers of people killed in Darfur. The psychological account presented here clearly has implications for motivating greater response to humanitarian crises other than genocide and certainly such implications should be pursued. I focus on genocide because it is a heinous practice, carried out by known human antagonists, that could in principle be stopped if only people cared to stop it. Apathy toward genocide and other forms of mass murder moves us closer to the loss of humanity.

African Union peacekeeping efforts in Darfur – another failure of the U. S. government to take meaningful action since September 2004 when Colin Powell returned from Sudan and labeled the atrocities there as “genocide.” We appropriately honor the one, Rosa Parks, but by turning away from the crisis in Darfur we are, implicitly, placing almost no value on the lives of millions there.

Acknowledgments Portions of this chapter appeared earlier in the paper “If I Look at the Mass I Shall Never Act: Psychic Numbing and Genocide,” that was published in *Judgment and Decision Making*, 2007, 2, 79–95. I wish to thank the William and Flora Hewlett Foundation and its President, Paul Brest, for support and encouragement in the research that has gone into this chapter. Additional support has been provided by the National Science Foundation through Grant SES-0649509.

References

- Associated Press. 1994, September 21. 38,000 shoes stand for loss in lethal year. In *The Register-Guard*. Eugene, OR: 6A.
- Baron, J. 1997. Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty* 14: 301–309.
- Barrett, L. F., and P. Salovey. 2002. *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence*. New York: Guilford Press.
- Bartels, D. M., and R. C. Burnett. 2006. Proportion dominance and mental representation: Construal of resources affects sensitivity to relative risk reduction. University of Chicago Personal Web Page. <http://home.uchicago.edu/~bartels/papers/Bartels-Burnett.pdf>. Accessed 24 November 2008.
- Bass, R. 1996. *The Book of Yaak*. New York: Houghton Mifflin.
- Batson, C. D. 1990. How social an animal? The human capacity for caring. *American Psychologist* 45: 336–346.
- Batson, C. D., K. O’Quin, J. Fultz, M. Vanderplas, and A. Isen. 1983. Self-reported distress and empathy and egoistic versus altruistic motivation for helping. *Journal of Personality and Social Psychology* 45: 706–718.
- BBC News, UK edition. 2005, November 19. Sparrow death mars record attempt. <http://news.bbc.co.uk/1/hi/world/europe/4450958.stm>.
- Clark, M. S., and S. T. Fiske. 1982. *Affect and Cognition: The Seventeenth Annual Carnegie Symposium on Cognition*. Hillsdale, NJ: Erlbaum.
- Coke, J. S., C. D. Batson, and K. McDavis. 1978. Empathic mediation of helping: A two-stage model. *Journal of Personality and Social Psychology* 36: 752–766.
- Daillaire, R. 2005. *Shake Hands with the Devil: The Failure of Humanity in Rwanda*. New York: Carrol & Graf.
- Dillard, A. 1999. *For the Time Being*. New York: Alfred A. Knopf.
- Dubinsky, Z. 2005. The lessons of genocide [Review of the book *Conspiracy to Murder: The Rwandan Genocide*]. *Essex Human Rights Review* 2: 112–117.
- Epstein, S. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49: 709–724.
- Eisenberg, N., and P. Miller. 1987. Empathy and prosocial behavior. *Psychological Bulletin* 101: 91–119.
- Fechner, G. T. 1860/1912. Elements of psychophysics. Classics in the History of Psychology. York University’s Classics in the History of Psychology. <http://psychclassics.yorku.ca/Fechner/>. Accessed 1 December 2008.
- Fenske, M. J., and J. E. Raymond. 2006. Affective influences of selective attention. *Current Directions in Psychological Science* 15: 312–316.

- Fetherstonhaugh, D., P. Slovic, S. M. Johnson, and J. Friedrich. 1997. Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty* 14: 283–300.
- Forgas, J. P. 2000. *Feeling and Thinking: The Role of Affect in Social Cognition*. Cambridge, UK: Cambridge University Press.
- Friedrich, J., P. Barnes, K. Chapin, I. Dawson, V. Garst, and D. Kerr. 1999. Psychophysical numbing: When lives are valued less as the lives at risk increase. *Journal of Consumer Psychology* 8: 277–299.
- Galanter, E. 1962. The direct measurement of utility and subjective probability. *American Journal of Psychology* 75: 208–220.
- Gilbert, D. T. 2002. Inferential correction. In *Heuristics and Biases*. T. Gilovich, D. Griffin, and D. Kahneman, eds., 167–184. New York: Cambridge University Press.
- Gilbert, D. T. 2006, July 10. We're hard-wired to ignore global warming. In *The Register-Guard*. Eugene, OR: A9.
- Glover, J. 2001. *Humanity: A Moral History of the Twentieth Century*. New Haven: Yale University Press.
- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.
- Hamilton, D. L., and S. J. Sherman. 1996. Perceiving persons and groups. *Psychological Review* 103: 336–355.
- Hume, D. 1960. *An Enquiry Concerning the Principles of Morals*. La Salle, IL: Open Court (Originally published in 1777).
- Jenni, K., and G. Loewenstein. 1997. Explaining the “identifiable victim effect”. *Journal of Risk and Uncertainty* 14: 235–257.
- Kahneman, D. 2003. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist* 58: 697–720.
- Kahneman, D., and Frederick, S. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics of Intuitive Judgment: Extensions and Applications*. T. Gilovich, D. Griffin, and D. Kahneman, eds., 49–81. New York: Cambridge University Press.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–291.
- Kingsolver, B. 1995. *High Tide in Tucson*. New York: HarperCollins.
- Kogut, T., and I. Ritov. 2005a. The “Identified Victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making* 18: 157–167.
- Kogut, T., and I. Ritov. 2005b. The singularity of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes* 97: 106–116.
- Lacey, M. 2005, May 18. The mournful math of Darfur: the dead don't add up. *The New York Times*: A4.
- Ledoux, J. E. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.
- Lifton, R. J. 1967. *Death in Life: Survivors of Hiroshima*. New York: Random House.
- Loewenstein, G., E. U. Weber, C. K. Hsee, and E. S. Welch. 2001. Risk as feelings. *Psychological Bulletin* 127: 267–286.
- Melvyn, L. 2004. *Conspiracy to Murder: The Rwandan Genocide*. London: Verso.
- Mowrer, O. H. 1960. *Learning Theory and Behavior*. New York: John Wiley & Sons.
- Neville, P. 2004, February 15. Statistics disguise a human face. In *The Register-Guard*. Eugene, OR.
- Power, S. 2003. *A Problem from Hell: America and the Age of Genocide*. New York: Harper Perennial.
- Schabas, W. 1999, January 7. The genocide convention at fifty (Special Report 41). United States Institute of Peace. <http://www.usip.org/pubs/specialreports/sr990107.html>. Accessed 3 December 2008.
- Schelling, T. C. 1968. The life you save may be your own. In *Problems in Public Expenditure Analysis*. S. B. Chase, Jr., ed., 127–176. Washington, DC: Brookings Institute.

- Schroeder, P., and Schroeder-Hildebrand D. 2004. *Six Million Paper Clips: The Making of a Children's Holocaust Museum*. Minneapolis, MN: Kar-Ben.
- Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2002. The affect heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment* T. Gilovich, D. Griffin, and D. Kahneman, eds., 397–420. New York: Cambridge University Press.
- Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24: 1–12.
- Slovic, S., and P. Slovic. 2004. Numbers and nerves: Toward an affective apprehension of environmental risk. *Whole Terrain* 13: 14–18.
- Small, D. A., and G. Loewenstein. 2003. Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty* 26: 5–16.
- Small, D. A., and G. Loewenstein. 2005. The devil you know: The effects of identifiability on punishment. *Journal of Behavioral Decision Making* 18: 311–318.
- Small, D. A., G. Loewenstein, and P. Slovic. 2007. Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes* 102: 143–153.
- Song, J. 2002, April 26. Every dog has its day – but at what price? In *The Register-Guard*. Eugene, OR: 1.
- Stanovich, K. E., and R. F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23: 645–726.
- Stanovich, K. E., and R. F. West. 2002. Individual differences in reasoning: Implications for the rationality debate? In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, D. W. Griffin, and D. Kahneman, eds., 421–444. New York: Cambridge University Press.
- Stevens, S. S. 1975. *Psychophysics*. New York: Wiley.
- Susskind, J., K. Maurer, V. Thakkar, D. L. Hamilton, and J. W. Sherman. 1999. Perceiving individuals and groups: Expectancies, dispositional inferences, and causal attributions. *Journal of Personality and Social Psychology* 76: 181–191.
- Tomkins, S. S. 1962. *Affect, Imagery, and Consciousness: Vol. 1. The Positive Affects*. New York: Springer.
- Tomkins, S. S. 1963. *Affect, Imagery, and Consciousness: Vol. 2. The Negative Affects*. New York: Springer.
- Ubel, P. A., J. Baron, and D. A. Asch. 2001. Preference for equity as a framing effect. *Medical Decision Making* 21: 180–189.
- Västfjäll, D., E. Peters, and P. Slovic. 2009. Representation, affect, and willingness-to-donate to children in need.
- Västfjäll, D., E. Peters, and P. Slovic. (Manuscript submitted for Publication). Compassion fatigue: Donations and affect are greatest for a single child in need.
- Vuilleumier, P., J. L. Armony, and R. J. Dolan. 2003. Reciprocal links between emotion and attention. In *Human Brain Function*. K. J. Friston, C. D. Frith, R. J. Dolan, C. Price, J. Ashburner, W. Penny, et al., eds., 419–444. New York: Academic Press.
- Weber, E. H. 1834. *De pulsu, resorptione, auditu et tactu*. Leipzig: Koehler.
- Zajonc, R. B. 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist* 35: 151–175.

Marketing Risk: Emotional Appeals Can Promote the Mindless Acceptance of Risk

Ross Buck and Whitney A. Davis

1 Introduction

Several contributors to this book maintain that it is important to take emotions into account in order to make rational decisions about the moral acceptability of technological risks: that we need emotions to judge whether a risk is morally acceptable. The present chapter addresses this question in terms of two ways of approaching and communicating risk: via rational argument and emotional appeal. At first glance, reason seems morally superior to emotion, because emotional appeals smack of manipulation. However, recent research illustrates that emotion is necessary for effective judgment in other realms, and this evidence is critical in understanding judgments about risk. The chapter discusses implications of emotional factors for understanding risk perception, suggesting that emotion has been largely overlooked in the design of warnings and that emotional considerations are necessary in the design of effective warnings. At the same time, advertisers and marketers have used emotion effectively, sometimes to encourage dangerous behavior and sometimes to promote the mindless acceptance of risk. These phenomena are explored with regard to how warnings function in everyday risks – diving risks, fire risks, and risks from alcohol and tobacco – with the assumption that these reveal principles important in responding to emerging technological risks.

2 Emotion and Reason in Persuasion and Risk Perception

2.1 *The High Road and Low Road to Cognition*

Dual modes of cognitive processing. Recent studies have recognized two different sorts of knowledge modes that have implications for the effectiveness of persuasion in general, and risk perception in particular (deTurck et al. 1993).

R. Buck (✉)

Communication Sciences, University of Connecticut, Storrs, CT, USA
e-mail: ross.buck@uconn.edu

In Chaiken's (1980, 1987; Chaiken and Eagly 1983) heuristic-systematic model, "systematic processing" involves accessing, scrutinizing, and integrating relevant information to reach a judgment, while "heuristic processing" involves the use of simple decision rules – cognitive heuristics – to reach a judgment. Similarly, Petty and Cacioppo's (1986) Elaboration Likelihood Model (ELM) distinguished between rational "central route" and "peripheral route" cognitive processing. Emotion was seen to be important in the peripheral route to persuasion where the issue at hand has relatively low involvement or personal relevance to the individual, and there is therefore little incentive to devote scarce cognitive resources to evaluating the arguments. The central route and Chaiken's systematic processing both demand and consume effortful and "mindful" analytic cognitive capacities. The theories differ in their conceptualizations of peripheral route versus heuristic processing, but in both cases such processing is regarded as less "mindful" and rational. Both of these approaches imply that the persuasion process may be influenced by emotion, but in both cases it is the judgment process that really counts, as it were: persuasion and risk perception per se would be based upon a central rational or "cold-cognitive" judgment process.

Emotion as syncretic cognition. Based upon neuropsychological theory and research, Tucker (1981) distinguished two sorts of cognition that mirror the central/systematic versus peripheral/heuristic distinctions in many respects. *Syncretic cognition* is "hot," direct, and immediate; while *analytic cognition* like central/systematic processing involves "cold," sequential, and linear information processing. In Tucker's view, emotion involves syncretic cognitive processing, and he related the distinction between analytic and syncretic cognition to processing modes characteristic of the left and right cerebral hemispheres, respectively.

Syncretic/emotional cognition involves memory and processing systems in the brain that are separate from those of analytic/systematic cognitive processing, are organized differently, and obey different rules (LeDoux 1994; Panksepp 1994). LeDoux (1996) distinguished a "high road" and "low road" to cognition, showing that emotion-related structures associated with the amygdala region of the brain receive input about events that is earlier than and potentially independent of input to relevant neocortical sensory systems associated with analytic processing. Furthermore, LeDoux outlined two central memory networks that operate simultaneously and in parallel: explicit or declarative memory which involves the hippocampus, and implicit or emotional memory which involves the amygdala (1994, p. 312). The cognitive attribution theorist Richard Lazarus (1991) acknowledged that LeDoux's findings and the distinction between analytic and syncretic knowledge effectively demonstrate that "raw" emotion indeed constitutes a kind of knowledge that can precede, and indeed contribute to, analytic knowledge: an "automatic mode of meaning generation" (Lazarus 1994, p. 215). The differentiation of analytic and syncretic cognition blurs the usual distinction between emotion and cognition: the subjective experience of emotion or affect becomes a *type of cognition*: a type of knowledge. In the present view, *affect* is defined as *the direct knowledge of feelings and desires, based upon readouts of specifiable neurochemical systems evolved by natural selection as phylogenetic*

adaptations functioning to inform the organism of bodily events important in self-regulation (Buck 1985, 1994, 1999). Human beings experience affects immediately and directly (*knowledge-by-acquaintance*); the phenomenological subjective reality of affect is self-evident.

The ARI model. Buck and Chaudhuri (1994) proposed a model of the interaction of emotion and reason in the context of advertising and persuasion, suggesting there are in effect two persuasion processes that take place simultaneously and interactively: a high road rational influence process involving analytic cognition that works much as central route/systematic processing theories in attitude change predict, and a low road emotional influence process involving a qualitatively different sort of cognition (Buck et al. 1995, 2004). In this Affect-Reason-Involvement (ARI) model, emotion and reason are considered to be qualitatively different kinds of cognitive systems which interact with one another (Buck 1999). Reason involves Tucker's linear and sequential analytic cognition, and LeDoux's high road involving declarative memory and the hippocampus. In contrast, emotion involves holistic and synthetic syncretic cognition, and LeDoux's low road, implicit/emotional memory, and the amygdala. Moreover, emotional and rational cognition are seen to be associated with two qualitatively different but simultaneous and interacting "streams" of communication: spontaneous and symbolic communication associated respectively with the right and left hemispheres (Buck 1984, 1988; Buck and Van Lear 2002).

The relationship between emotion and reason is expressed by Fig. 1. This represents an interaction of rational and emotional systems with the relative importance of reason increasing as one goes from the left to the right. That is, in this representation, the influence of emotion is always present, while the influence of reason increases from zero on the left to a high value relative to emotion on the right. On the extreme left of the affect/reason continuum (*A/R Continuum*), the influence of affect is total: reason has no influence. As one goes to the right, reason exerts an increasing influence relative to affect, but the influence of affect never falls to zero.

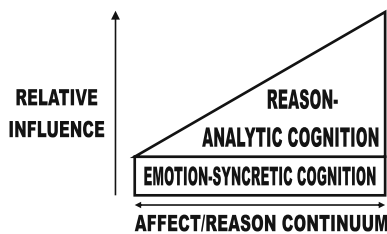


Fig. 1 The suggested relationship of emotion and reason (syncretic and analytic cognition). At the *left*, the influence of emotion is absolute, with reason increasing in influence toward the *right*. At the right extreme, reason is dominant but emotion still exerts influence

The position of an object or message on the A/R continuum is determined by the amount of feeling relative to thinking reported by an individual toward that object, reflecting the ratio of affect to reason in that object, or in the message advocating for or against that object (Chaudhuri and Buck 1993). Thus, an affectively loaded object

or message has a high A/R ratio, while an object or message to be dealt with “mindfully” has a low A/R ratio. For example, participants might be asked how much they feel specific emotions about a given technological risk (secure, afraid, angry, trusting) and how much they know and think about it. For consumer products, candy and snack foods have high A/R ratios; automobiles, computers, airline services, and paper products have relatively balanced in A/R scores; and appliances, insurance policies, and laundry products have low A/R ratios. There does not need to be a match between the A/R ratio of the object and a message advocating it: for example insurance policies are rated to be rational objects, but a major insurance company in the United States has an animated gecko as its spokesperson. The adorable amphibian has no relationship with insurance except a clang association with the company name. Indeed, the amusing source of the advertising message may function to infuse a relatively boring and uninteresting object with emotional appeal.

Social and moral emotions. Theories of emotion often distinguish between emotion dimensions, such as strong-weak and pleasant-unpleasant; and specific emotion types, such as the primary affects of happiness, sadness, fear and anger. These approaches have tended to emphasize “individualistic” emotions associated with individual survival, and overlook sex and other “prosocial” emotions associated with species survival posited in MacLean’s (1993) triune theory of the brain. Recent studies have supported MacLean’s approach, in that they have focused upon neurobiological systems that regulate social organization involving attachment and bonding (see Buck 1999; Carter et al. 1997; Panksepp 1993; Kosfeld et al. 2005). The subjectively experienced affects associated with these systems appear to involve specific neurochemicals including serotonin, oxytocin, gonadotropin releasing hormone, and the endorphins. These participate in the regulation of a vast array of positive social behaviors involving feelings of erotic arousal, nurturance, intimacy, caring, trust, and love.

Challenges to the strong feelings associated with these biologically-based attachment systems are arguably the biological foundation of “higher level” social and moral emotions (see Buck 1988, 1999). These include the social emotions of pride/hubris, shame/guilt, envy/jealousy, and pity/scorn; and negative and positive moral emotions including resentment, humiliation, indignation, and contempt versus gratitude, respect, elevation, admiration, and trust (Buck 2004). Understanding the role of emotions in persuasion in general and risk perception in particular must take prosocial and higher-level social and moral emotions into account.

2.2 Implications for Persuasion and Risk Perception

The effectiveness of warnings. A “warning” may be defined as a message that identifies prohibited behavior, and communicates the risk and consequences of such behavior in a manner reasonably calculated to deter such behavior. In contrast, an “instruction” identifies prohibited behavior without communicating risks or consequences of engaging in that behavior, such as a speed limit sign on a roadway.

The effectiveness of warnings in deterring unsafe behavior has been the subject of considerable research, but substantial disagreement remains. For example, McGrath and Downs (1992) argued that “warning labels are often an ineffective method for communicating safety information” (p. 24), and that the proliferation of warning labels is counterproductive.

Many studies of warning effectiveness have suggested that information processing occurs in successive stages (deTurck et al. 1995). For example, Wogalter and Laugherty (1996) proposed successive stages of attention, comprehension, persuasion to ensure correct attitudes and beliefs, and motivation to produce the desired behavior. The factors leading to effective persuasion and motivation are thus critical to effectiveness, in addition to comprehensibility and legibility. These are enhanced by clear statements of the hazard and consequences of the dangerous behavior, as well as instructions about how to avoid the hazard (Wogalter et al. 1987).

In a meta-analysis of studies of the effectiveness of warnings which included a no-warning control condition, Cox et al. (1997) concluded that, on the average, warnings do indeed increase safe behavior, albeit to a modest extent. However, they also found considerable variability, such that the beneficial effect of some warnings was large, some warnings had no effect, and some warnings actually appeared to have a boomerang effect, being associated with more unsafe behavior than the no-warning control condition. Clearly, there are substantial differences in the effectiveness of warnings, and the bases of these differences are poorly understood.

The American National Standards Institute (ANSI) approved and published general voluntary standards for warning labels in 1991. Revised in 1998 and 2002 these ANSI standards were designed to be “recognizable, legible, readable, and understandable, and therefore, more effective” (Martin and Deppa 1997, p. 821). The warning labels use a two- or three-panel format. The top panel contains a safety alert symbol (a triangle surrounding an exclamation point) and signal word with background color indicating the level of risk (DANGER/red, WARNING/orange, and CAUTION/yellow). The bottom message panel includes text that describes the hazard. The middle panel, which is optional, contains an iconic pictorial “international symbol” that graphically represents the hazard and, sometimes, its consequences. The standards recommend that the pictorial symbol not be used without text unless it has been validated, with the criterion of validation being legibility and comprehensibility (Deppa and Martin 1997).

Emotion in warnings. Certainly, intelligibility and comprehensibility are necessary to the effectiveness of warnings, but they are not sufficient. Hazard and danger by their very nature involve highly affective “fight-or-flight” processes, and emotions such as fear and anxiety, as well positive emotions involving happiness, relaxation, and security; are clearly involved in avoiding harm. However, although Zuckerman and Chaiken (1998) applied Chaiken’s (1980, 1987) heuristic-systematic model to explain the effectiveness of product warning labels, in general the scientific literature on warning effectiveness has rarely mentioned emotion explicitly.

We suggest that, to be effective – that is, to be persuasive and to motivate behavior– warnings must not only present a legible, comprehensible message with

clear consequences and instructions, they must also turn on the “low road” to cognition: they must activate the amygdala and implicit/emotional memory systems. More specifically, we argue that, to be effective, warnings must (a) command attention, (b) galvanize memory, and (c) evoke emotion. In particular, the ability of a label to communicate the risks of the prohibited behavior on an emotional level is not only important, but vital, to the successful communication of danger to non-English readers and children under 12 years of age who lack the ability to fully draw implications from a verbal message. The use of pictorial international symbols that portray consequences may incidentally increase the emotional content of warning messages. Many such symbols are intrinsically emotional. However, such effects are unintentional, and emotional factors are not part of the process explicitly considered in the design of such symbols.

Advertising and emotion. Although the characteristics of being noticeable, memorable, and emotional do not always apply to signs intended to be instructions and warnings, they are almost always apparent in the design of advertisements. Advertising and marketing professionals have long appreciated the importance of emotion. Inspection of a few television commercials or magazine advertisements is enough to convince an objective observer that advertising and marketing professionals are well aware of the importance of emotion in the practical realm of persuasion and influence. These include advertisements that can encourage *unsafe* behavior: the tobacco industry and alcohol industries spend millions on advertising to promote the use of potentially dangerous products. Even more deviously, advertisements and promotional messages can have the opposite effects of effective warnings: turning attention away from potential danger and instilling memories and emotions that are incompatible with recognizing danger and risk. Such messages can in effect induce the mindless acceptance of risk.

2.3 Cultural, Individual/Situational, and Perceptual Factors in Dangerous Behavior

The effectiveness or ineffectiveness of warnings must be considered in the context of other factors important in the determination of dangerous behaviors. In most cases, accidents are multi determined, resulting from a confluence of factors, some of which are not always immediately apparent. Behaviors in dangerous situations occur in the context of three levels of influence. *Cultural* influences involve shared expectations about the danger or lack of danger in a behavior. These expectations make up cultural norms regarding the behavior. As noted, advertisements can often create shared expectations or norms that mitigate the communication of danger. *Individual/situational* influences involve the actions of the individual in a given situation, and bring in questions about the extent that the dangerous behaviors of individuals are based upon their own characteristics (such as “risk-taking” personality patterns) or rather, have situational influences as their predominant source. *Perceptual/ecological* factors involve whether the danger is immediately apparent

at the point of performing the behavior: that is, whether the warning has ecological presence such that the individual can immediately perceive the danger, as he/she is about to act.

The next section of this chapter considers the role of warnings in communicating the dangers of diving, mattress safety, and the marketing of alcohol and tobacco. The chapter considers cultural, individual/situational, and perceptual factors involved, and the role of emotional factors in mitigating risk.

3 Diving-Related Injuries and Warnings

Diving accidents are all too common, and often result in devastating spinal injuries causing paraplegia and quadriplegia. Gabrielsen and Spivey (1990) estimated that diving accidents account for over 10% of spinal cord injuries, with pool diving accounting for between 150 and 250 such incidents each year in the United States. Diving accidents remain, by far, the leading cause of sports-related spinal injuries. The projected costs for lifetime medical and attendant care for an 11 year-old C-4 Frankel-Class quadriplegic exceeds \$27,000,000.

3.1 A Historical Perspective

Gabrielsen and Spivey (1990) reported that the 48 in. deep aboveground pool industry began in the United States in the late 1950s. A pool industry organization, the National Swimming Pool Institute (NSPI), was formed in 1956. Their standards, first published in 1958 and in 1961 for residential pools, omitted any mention of a need for warnings. Although pool-related diving injuries began to appear in the early 1960s, the revisions of standards published by NSPI in 1969, 1972, and 1974 again failed to mention warnings despite the emerging realization of dangers, with reports of diving injuries resulting in quadriplegia appearing in industry publications. By 1975, thousands of people had been paralyzed by diving injuries, and individual pool builders and manufacturers began to take the initiative to produce decals, but most of these merely consisted of “No Diving” instructions rather than warnings. Advertisements depicting children and adults diving into these 48 in. deep pools were deployed in 1963, but discontinued by 1971 due to the steadily- increasing number of catastrophic diving injuries that resulted from this product’s adoption by ordinary consumers. It was not until 1980 that NSPI published official standards, which merely called for a “No Diving” instruction placard with lettering “not less than ¼ in. (0.6 cm) in height.” ANSI standards suggest that this font size is virtually unreadable at distances of more than 9 ft.

In 1977, a major pool retailer directed the manufacturer of the pools it sold to place on the pools a label that said: DANGER – NO DIVING – SHALLOW WATER- DIVING CAN CAUSE PARALYSIS. It also included an international symbol showing a diver covered by a slant line within a circle. However, the first

comprehensive studies of pool signage did not come until the late 1980s: a focus group study contracted by NSPI, and a study conducted by the American Institute for Research (AIR) contracted by the Consumer Products Safety Commission (CPSC). The latter study concluded that warning signs could be effective, and that each should have a signal word, written text, and a “pictorial which graphically portrayed the prohibited action and its consequences” (Gabrielsen and Spivey 1990, Chapter 15, p. 3).

An example of what would be expected to be a more effective warning based on the AIR research shows the symbol of the diver striking his head with attendant lightning bolts communicating the prohibited conduct, illustrating in an emotionally loaded manner the prospect that a diver can be hurt. Moreover, a separation between the diver’s head and body graphically illustrates the consequences of contacting the pool bottom: a broken neck and paralysis. The top panel contained the safety alert symbol and WARNING/orange signal word, and the lower message panel included text that clearly and accurately described the hazard.

3.2 Cultural Factors: Early Pool Advertisements

The above-ground pool product was originally marketed to an audience seeking a portable and inexpensive alternative to the in-ground pool. Luxury models were also designed for those of means, but whose yard conditions would not permit the excavation necessary for a pool (i.e. high water table or mountainside conditions). The market demographic was almost invariably dominated by young parents with two or more children.

As noted, early advertisements created by the aboveground pool industry to promote their products often depicted diving into 48 in. deep aboveground pools and other dangerous behaviors. The CEO of the largest manufacturer of such pools testified that they and other manufacturers used such advertisements starting in 1963. It is apparent by these materials and the CEO’s testimony about how they were used, that a “culture” was created in early adopters of 48 in. deep aboveground pools that encouraged diving among other dangerous behaviors. Such a culture is composed of shared expectations that diving into 48 in. deep aboveground pools on the part of children is anticipated, expected and indeed tacitly encouraged.

One of the basic findings in social psychology is that, once started, a culture tends to endure through social influence processes involving learning, modeling, and imitation; even when the original motivating factors are removed. For example, the classic study by Sherif (1965) used the autokinetic effect, the tendency to perceive movement in a single stationary light in an otherwise dark room. The apparent movement comes from eye movements. People experiencing this phenomenon judging the extent of movement by speaking aloud typically begin with variable judgments, but then settle on a single “individual norm” of, say, 9 in. for one person, 3 in. for another. If two people who have arrived at individual norms make audible judgments in each other’s presence, Sherif found that they influence

one another, arriving at a “group norm” that is usually a compromise between the individual judgments, say 5 in. To test this influence process, Sherif created three-person groups in which, unknown to the others, one person was instructed to give extreme judgments. That person was found to influence the other two, so that they, too, began to give extreme judgments. The first person then left and was replaced by a new, naïve person. Even though the original person had left, the extreme judgments of the remaining two influenced the new person. Then another of the original persons left, and a new naïve person was influenced, and so on. The extreme judgments of the original person lasted through many such “generations,” despite the fact that the originator of the extreme judgments was long gone. A group norm had been established, and was perpetuated.

Analogously, the influence of advertisements depicting children diving into 48 in. deep aboveground pools lived on long after the advertisements have been withdrawn. These images reinforced what people want to do anyway: it is fun to dive, the potential danger of diving into 48 in. deep pools is not readily apparent, and many people including adults dive repeatedly without injury. The advertising images of diving children constitute models in the sense used by Bandura and Walters (1963): that is, the images implicitly endorsed, and actively encouraged the imitation of, the behavior depicted just as violent images in media can endorse and encourage aggressive behavior. This culture was actively stimulated by the pool industry via attractive and colorful advertisements for 9 years, and based upon the above reasoning these advertisements also programmed later adopters and users of the product. That is, the programmed acceptability of diving into 48 in. deep aboveground pools had been passed down orally and by example from one generation to the next. Even though a given individual might not necessarily see these advertisements per se, they could nevertheless influence his/her expectations and behavior.

The pool industry created this culture through advertising, and they profited handsomely from it. But the culture started by these advertisements persists even today. The legacy of the early advertisements is like a genie let out of the bottle: once images of children diving into 48 in. deep pools were widely published and disseminated, they cannot be taken back. The implicit lesson is, diving is cool and exciting, while rules are dull and boring. This lesson influences caregivers as well as children: rules against diving may be given lip service, but they often are not really taken seriously. They constitute instructions rather than warnings: “Walk the dog.” “Clean up your room.” “Take out the garbage.” “Don’t dive.”

3.3 Individual/Situational Factors: “Risk Taking,” Play, and Natural Exploration

“*Risk taking.*” Dangerous behavior always involves an interaction between characteristics of the individual and those of the situation. Some kinds of risky behaviors – sky diving, rock climbing, car racing – are clearly dangerous and appeal to a minority of individuals often characterized as “risk-takers” (Zuckerman 2007).

Such characterizations apply only to adults, because the notion of “taking a risk” implies that the individual is capable of an adult level of cognitive functioning that can appreciate the nature of the risk, that is, appreciate logically the long-term consequences of dangerous behavior.

In contrast, children under 12 have not attained a formal operational level of thinking that makes this possible: there is a cognitive gap between the context dependency and concrete operational functioning of a child and the ability to appreciate risk. Indeed, a thorough literature search on *PsychInfo* back to 1887 revealed that there are virtually no studies of “risk taking” or “sensation seeking” among children less than age 12.

Children do however differ in their general fearfulness, and this can have consequences for their natural tendency to engage in dangerous behavior. Cook et al. (1999) asked 130 fourth graders to report on their level of excitement versus fear in response to common play situations, including situations involving water play. A week later, the same children were observed at a public swimming pool. Children’s reports of experiencing fear were related negatively to rates of actual dangerous behavior and positively to rates of protective behavior at the pool. Thus, children’s perceptions of their own fearful emotional reactions predicted their tendencies to engage in dangerous pool behaviors. This makes it all the more important that bolder children be warned of the potentially catastrophic consequences of diving injury.

Play. Although individual differences in fearfulness may have some effect upon dangerous behavior in children, the pool-play situation itself is arguably more important. Children, and adults, do not typically engage in a great deal of rational, analytic, systematic thinking during play at a pool: that is, the control of their behavior tends to be on the left side of the A/R continuum illustrated in Fig. 1. The pool-play environment itself tends to “pull out” exuberant, high-spirited, and passionate behaviors from everyone involved, regardless of individual differences in personality, fearfulness, or anything else. Such behavior can easily become rambunctious, rowdy, and potentially dangerous, and this can of course be exacerbated by the use of drugs and alcohol. It can also be exacerbated by natural tendencies of children to “push the envelope” in their natural exploration of their environment.

Natural exploration. Children become competent in the context of exploratory play. They naturally try new things when they feel capable. Piaget (1971) conceptualized a process of cognitive growth cycles in terms of assimilation and accommodation: a child faced with an experience that is assimilable but not completely accommodated will be intrinsically motivated to take that experience on and may become completely immersed. However, once the experience is accommodated, the child will lose interest and go on to something else. White (1959) reported an incident where a little girl was so immersed by a puzzle that she never noticed when the teacher lifted her, chair and all, onto a table. She continued to play for 20 min, and then looked up and realized where she was. But then the little girl never touched that puzzle again. It had served its function of cognitive growth, and now was boring: she went on to new experiences.

Such behavior does not constitute risk-taking, it is exploration that is fundamental to cognitive development and growth. Nevertheless, such behavior can be

dangerous. In one tragic case, an 11-year old girl who had been doing repeated shallow dives into a 48 in. deep pool while playing with her friends, decided to try a racing dive from a bench located by the poolside. She misjudged, hit the bottom, and was rendered quadriplegic in an instant.

3.4 Perceptual/Ecological Factors

At the moment she began her last dive, this young girl could not perceive the danger concealed in the pool. Although 48 in. deep pools are too shallow for diving, this is not immediately apparent. These pools lack effective cues to their shallowness: most aboveground pools have a light blue vinyl liner as the water container, which interferes with the perception of depth. The concealed dangers of diving can be, and often have been, missed even by experienced adults. Gabrielsen and Spivey (1990) have advocated the placement of black lines on the bottom of aboveground pools to aid the perception of depth, but this is rarely done.

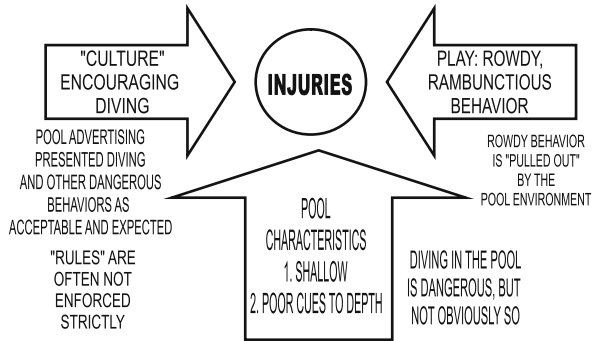
Every parent associated with the girl would later admit that they dove in these very pool models when they were 12 years of age, and saw no real risk. The girl entered the water merely two degrees steeper than she should have, and she broke her neck at impact: few realize that a two-degree variation in the dive angle results in hydro-dynamic rotation of the upper body in a water-column. The water column helps accelerate the diving body, and at the same time, orients the skull for a neck-breaking impact. It does not readily occur to most that a diving mistake brings with it paralysis or death. There is even less comprehension of the personal loss associated with six more decades in a wheelchair. Even if some risk is comprehended, the swamping effect of the situational excitement of play serves to drown-out any residual low-road or high-road message reception.

Because diving dangers are not readily apparent, and because low-road and high-road message reception is situationally compromised in both the diver and many supervising adults, it is critical that an effective warning be presented in the actual pool environment, in a manner that naturally captures attention, galvanizes memory, and evokes emotion. It requires a stern, memorable, ugly, unpleasant, emotionally evocative and graphically explanatory message to cause pool users – particularly children – to pause, reflect, and reconsider unsafe behavior. As Gabrielsen and Spivey (1990) put it, “signs located at the place where diving is prohibited represent the last opportunity to communicate with the person who is contemplating a dive. Information properly communicated could cause the individual to alter the intention to dive” (Chapter 15, p. 3).

3.5 Conclusions

The three suggested levels of influence involved in dangerous behavior in the case of diving accidents are summarized in Fig. 2. Those in the pool industry had and

Fig. 2 Cultural, individual/situational, and perceptual/ecological influences often combine in risky behaviors and accidents: here is an example from diving risk



have a moral obligation to provide an effective brake on the unsafe behavior that they themselves encouraged. There is a wider obligation to assure that children’s environments are safe for normal play and exploration, or if not, to provide effective warnings. The 48 in. deep aboveground pool does not offer a safe environment for children. The dangers are grave and barely perceptible. The risk is certain, but lurks beneath the din of recreational excitement. Accordingly, there is an obligation to provide effective warnings at these pools. The next section of this chapter considers the nature of effective warnings in the context of mattress safety.

4 Mattress Safety: Polyurethane Foam and Fire Danger

To the average consumer, mattresses are not considered to be particularly hazardous, and cultural and individual factors are not greatly relevant to the risks posed by mattresses. The major factor of concern is perceptual: the mattress does not appear to be dangerous. But there is hidden, yet passive, risk.

Mattresses ignite from heated appliance cords, child play with matches, or as fires secondary to another item burning. Due to the combustive strength and toxicity of its component products, however, this product turns an otherwise survivable fire into a conflagration.

The polyurethane foam in conventional innerspring mattresses sold in the United States prior to 2007 can be highly combustible. Some constituent components have the combustibility of kerosene and gasoline. The fuel load in the foam burns very rapidly, and can cause an explosive “flashover” in 3–5 min from ignition. Also, when the foam smolders or ignites, it emits smoke containing carbon monoxide, carbon dioxide, and cyanide gas. Cyanide gas is a nerve agent that disables the bed fire victim while they burn. Among the hazardous ingredients in polyurethane foam is isocyanate, a cyanide derivative which was responsible for killing and injuring thousands in the Union Carbide pesticide plant disaster in Bhopal, India in 1984 (Eckerman 2004). The dangers of polyurethane were also demonstrated in a 2003 fire in a rock club in Westerly, Rhode Island, which killed 100 and injured 200.

Polyurethane foam used as soundproofing was set alight by band pyrotechnics. The building was fully engaged by fire within 3 min. Several exits were available, but victims were killed by “smoke inhalation.” It is perhaps more precise to say that they were felled by poison gas.

In 1993, polyurethane foam manufacturers began issuing warnings to mattress manufacturers regarding potentially fatal hazards. The 8.5 in. by 11 in. label stated:

WARNING!

**FLAMMABLE POLYURETHANE FOAM
FOAM BURNS RAPIDLY**

When ignited, this foam burns rapidly resulting in
great heat, generating dangerous and potentially
toxic gas and thick smoke, consuming oxygen.
Burning foam can be harmful or fatal.

Keep foam away from open flames, sparks or
other heat sources, Do not smoke near this foam.

**IF FOAM STARTS BURNING
GET OUT!**

These warnings should be passed on to the ultimate users.

However, ultimate users have not been informed of this potentially fatal hazard. In 1971 California mattresses bore a small 2 in. by 3 in. label stating:

NOTICE: THIS MATTRESS HAS BEEN DESIGNED TO RESIST
COMBUSTION WHICH MAY RESULT FROM A SMOLDERING
CIGARETTE. THIS PRODUCT CONTAINS
NON-FLAME-RETARDANT POLYURETHANE FOAM.
AVOID CONTACT WITH OPEN FLAME.

Many requirements in other states offer even less information, and the mattress industry has criticized even this mild statement. This is the same industry whose members depict their mattresses in advertisements romantically surrounded by lit candles.

This example illustrates the principle that corporations often minimize and hide real risk in the pursuit of profit. The description “non-flame-retardant” is a confusing double-negative way to express that the foam burns rapidly, resulting in great heat, generating thick smoke, dangerous and potentially toxic gas including cyanide, and consuming oxygen. This constitutes deliberate obfuscation, not warning. Messages that deliberately misrepresent risky products as safe, thus undermining true warnings, have been termed *anti-warnings* (Boheme and Egilman, 2006), and these have been particularly prevalent in the advertising of tobacco products, as we shall see.

In February 2006 the Consumer Product Safety Commission (CPSC) approved a new flammability standard for mattresses made and imported into the United States. The new standards were designed to reduce the number of mattress fires from open-flame sources such as candles and lighters, and limit the amount of heat released in a mattress fire. The standards were implemented after July 1, 2007, although they

did not apply to mattresses in the inventory at that time. The CPSC estimated that the new rule could prevent as many as 270 deaths a year.

5 Alcohol and Tobacco Warnings

The major points of the present analysis of diving and mattress warnings apply to warnings regarding alcohol and tobacco as well. The small informational labels now appearing on cigarette and alcohol packaging in the United States are widely known, but they do not constitute warnings. Although the labels do state consequences of alcohol misuse and tobacco use, they arguably do not command attention or evoke emotion. For example, since 1984 tobacco labels have contained factual statements in small black and white type (e.g., “Surgeon General’s Warning: Quitting smoking now greatly reduces serious risks to you health”) that tend to be lost in the colorful, and attractive cigarette packs with large typeface brand names that have been heavily advertised (Fischer, Richards, Berman and Krugman, 1989). It is therefore not surprising that their effectiveness is modest at best, compared with efforts in other nations: tobacco messages on American cigarette packages are among the least effective in the world, and they are among the least emotionally arousing.

5.1 Cultural and Perceptual/Ecological Factors in Tobacco use and Alcohol Abuse

We saw that pool advertising showing images of children diving into 48 in. deep pools encouraged the impression that diving is fun and exciting, and fostered shared expectations that diving into such pools on the part of children is anticipated, expected and tacitly encouraged. These influences have lingered long after the advertisements themselves have been withdrawn. A similar case can be made with regard to alcohol and tobacco advertising, albeit on a vastly wider, global scale. The advertising of these potentially addictive and dangerous products has been spectacularly successful, and expectations about them favorable to the industries have become widely shared as a consequence. A notorious example of such advertising used a cartoon character, Joe Camel, to promote positive emotions toward smoking among children.

The efforts of anti-drunk driving and antismoking campaigns have made some headway as the dangers inherent in misuse of alcohol and use of tobacco have become more widely known and appreciated, but it has been an uphill struggle against well-designed, expensive, and highly emotional advertising appeals. In 2005 the World Health Organization encouraged the use of messages employing larger color images on cigarette packs graphically illustrating the dangers of tobacco, such as a premature infant or damaged heart tissue. The images are large, memorable, ugly, unpleasant, emotionally evocative and graphically explanatory, and therefore are more likely to constitute effective warnings from the present point of view.

Tobacco companies in many countries have resisted following the WHO recommendations, arguing that larger warnings would violate commercial speech rights. In India, for example, pictorial warnings have been delayed by industry lobbying for nearly 7 years (Ramakant 2008). At issue are whether the larger and graphic messages are sufficiently effective to justify the burdens placed on the industry.

Canada has employed graphic and unpleasant images covering 30% of the cigarette pack since 2000, and there are proposals to enlarge them to 60%. Research has indicated that compared to American labels these produce stronger negative affective responses to smoking among both smokers and nonsmokers (Givel 2007; MacKinnon and Nohre 2006; Peters et al. 2007). Importantly, the Canadian labels did not produce defensive responses among smokers, as fear appeals may do. Also most smokers and non-smokers endorsed the use of Canadian-style labels in the United States (Peters et al. 2007). The International Tobacco Control Four Country Study is a longitudinal study employing four waves of surveys taken in 2002–2005. This compared 15,000 smokers from nations using labels that were well below the WHO standard (the United States and United Kingdom at baseline), slightly below the standard (Australia), enhanced to the standard (UK at follow-up), and at the standard (Canada). Results indicated that the more emotional labels produced higher levels of awareness and perceived effectiveness: the more emotional, the more effective (Hammond et al. 2007).

5.2 Individual/Situational Factors in Tobacco use and Alcohol Abuse

By law, alcohol and tobacco marketing cannot any longer be targeted at children in the United States (although the cultural influence of Joe Camel will no doubt long endure). However, natural adolescent curiosity and exploration are significant factors in tobacco and alcohol use among young people, not to mention dangerous illicit drug use and risky sexual behaviors. As with diving, the emotions and reasoning of children and adolescents must be taken into account: dangerous behaviors, and with older adolescents outright risk-taking, must be expected and acknowledged. As with the pool industry in the case of diving, the alcohol and tobacco industries have an obligation to provide effective brakes on unsafe behaviors that they themselves have encouraged.

The tobacco industry faces the inconvenient reality that it kills its best customers, so that inducing young people to smoke is a continuing challenge. The industry can no longer argue that smoking is safe or even healthy, as was done by advertising in the past. A legal settlement between states in the US and tobacco companies in 1998 prohibited the latter from taking “any action, directly or indirectly, to target youth. . . in the advertising, promotion or marketing of tobacco products.” But still, in every way possible, the industry seeks to associate tobacco with youth, health, thinness, attractiveness, sociability, sex, and power. R.J. Reynolds – the company that introduced the cartoon character Joe Camel – marketed flavored

cigarettes, including “Kauai Kolada,” a pineapple and coconut-flavored cigarette; and “Twista Lime” a citrus-flavored cigarette. In 2004, they introduced Camel flavors including “Winter Warm Toffee” and “Winter MochaMint.” Also, Brown and Williamson introduced flavored versions of Kool cigarettes; and the US Smokeless Tobacco Company marketed flavored chewing tobacco. Indeed, according to campaignfortobaccofreekids.org, the tobacco industry spends over \$12.4 billion per year to market smoking in the US

With increasing restrictions in advertising such marketing must be done less directly, though sponsorship of glamorous sports enterprises such as Formula One racing and NASCAR, rock concerts, and placement of smoking scenes in motion pictures and video games. Even overtly “anti-smoking” messages and advertisements can be manipulated: in one example, the rational manifest content of advertisements created by tobacco companies as part of settling product liability lawsuits seemed to be strongly anti-smoking and was accepted as such. However, research found that the actual effect of the messages was to increase smoking among young people, perhaps by presenting as affective latent content the message “only cool, mature, independent kids smoke.”

5.3 Implications for Understanding the Role of Emotion in Warnings

Understanding the role of emotion in tobacco messages relates to the more general issue of understanding the role of emotion in warnings. In a debate about the role of fear messages in tobacco control, Hastings and MacFayden (2002) argued that tobacco companies have succeeded through advertising in building long term attachment relationships with their customers, and suggested that to succeed anti-tobacco efforts must beat the industry at its own game, building antismoking “brands” which are trusted and respected by potential smokers. They cite the success of TheTruth.com campaign that is directed at uncovering the lies and manipulative tactics of the tobacco industry and highlighting its health effects in creative and amusing ways. They further suggest that fear messages lack this relational dimension, attempting simply to intimidate the errant smoker.

Biener and Taylor (2002) countered this criticism by citing the success of anti-tobacco advertisements in Massachusetts which presented victims of smoking in emotionally evocative ways. Judges rated the advertisements on positive emotions (entertaining, happy, funny); negative emotions (sad, frightening, disturbing); and whether the ad was believable and thought-provoking. Ads arousing negative emotions were rated as more thought-provoking and emotionally arousing than those arousing positive feelings. Surveys indicated that the negative ads were also recalled as being the most effective. Biener and Taylor also noted that these advertisements did not simply elicit fear, but also sadness at the grief of victims whose loved ones died, anger at tobacco companies for their pursuit of profit at any cost, and empathy and hope for smokers struggling to quit smoking (Beiner et al. 2000).

It is possible that these two relatively successful campaigns may actually be touching a similar cord in the audience. Both may elicit in different ways strong moral emotions by depicting smokers as deliberate victims of exploitation by the tobacco industry. The knowledge of the lies and manipulative practices of the industry can elicit feelings of humiliation and powerlessness in those who have yielded to the temptation of smoking. Unlike fear that motivates escape and avoidance behaviors, these moral emotions stir feelings of resentment and indignation that can motivate anger and corrective action against the industry. Fear is useful and necessary in warnings, motivating avoidance behavior in the immediate ecological presence of risk, but moral emotions can have a more general function of motivating openness to the truth and demanding fairness and an end to propaganda and exploitation in persuasive campaigns. At the least, these campaigns may directly target the aims of tobacco advertising to foster attachment relationships with their customers.

6 Conclusions

In conclusion, an adequate and effective warning needs to communicate the risk and consequences of the prohibited behavior on an emotional level as well as on a rational level, with due regard to the cognitive and linguistic abilities of various age and cultural groups exposed to the warning. Such warnings can be effective in reducing injuries, and although the reasons for the variability in the effectiveness of warnings noted previously is not known, it is possible that a key factor is the emotional content of the warning. Effective warnings act as “brakes” to stop dangerous behaviors. To be effective, warnings must command attention, galvanize memory, and evoke emotion. This allows the full activation and utilization of human cognitive abilities: syncretic/emotional abilities involving implicit memory systems as well as analytic/rational abilities involving declarative memory systems.

Moreover, while the evidence that emotional factors can guide good risk evaluation is compelling; there is also compelling evidence that emotional factors can disrupt good risk evaluation, and that such factors have been used to promote the mindless acceptance of risk. These principles are thoroughly understood in the advertising industry, and they can be and have been used in the exploitation of people for economic ends, regardless of safety. The tobacco industry has been proven to have engaged in blatantly shameless exploitation, promoting its products for decades while knowing of their addictive qualities and deadly effects, and deliberately hiding that knowledge, while spending millions to present an emotional image of smoking as exciting, youthful, healthy, and sexy, intentionally manipulating even children.

An important implication of this analysis is that the scientific evaluation and design of warnings must take lessons from the approach used by the advertising industry. Warnings should be evaluated for ecological presence and emotional

impact, and effects on attention and implicit memory, as well as for legibility and comprehensibility. Arguably, “human factors” research has often overlooked important emotional factors in the control of human behavior. Emotion research can enhance our understanding both of how warnings can be made to be more effective in communicating danger, and how emotional manipulation can undermine good risk perception in the pursuit of profit.

References

- Bandura, A., and R., Walters. 1963. *Social Learning and Personality Development*. New York: Holt, Rinehart & Winston.
- Beiner, L., G., McCallum-Keeler, and A. L., Nyman. 2000. Adult’s response to Massachusetts anti-tobacco television advertisements: Impact of viewer and advertisement characteristics. *Tobacco Control* 9: 401–407.
- Beiner, L., and T. M., Taylor. 2002. The continuing importance of emotion in tobacco control media campaigns: A response to Hastings and McFadyen. *Tobacco Control* 11: 75–77.
- Bohme, S. R., and D., Egilman. 2006. Consider the source: Warnings and anti-warnings in the tobacco, beryllium, and pharmaceutical industries. In *Handbook of Warnings*. M. S. Wogalter, ed., 635–644, Mahwah: Lawrence Erlbaum Associates. xxi, 841 pp.
- Buck, R., and C. A., Van Lear. 2002. Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication* 52: 522–541.
- Buck, R. 1984. *The Communication of Emotion*. New York: Guilford Press.
- Buck, R. 1985. Prime theory: An integrated view of motivation and emotion. *Psychological Review* 92: 389–413.
- Buck, R. 1988. *Human Motivation and Emotion*. New York: Wiley.
- Buck, R. 1994. Social and emotional functions in facial expression and communication: The readout hypothesis. *Biological Psychology* 38: 95–115.
- Buck, R. 1999. The biological affects: A typology. *Psychological Review* 106: 301–336.
- Buck, R. 2004. The gratitude of exchange and the gratitude of caring: A developmental-interactionist perspective of moral emotion. In *The Psychology of Gratitude*. R. A. Emmons, and M. McCullough, eds., 100–122, New York: Oxford University Press.
- Buck, R., and A., Chaudhuri. 1994. Affect, reason, and involvement in persuasion: The ARI model. In *Konsumenten Forschung. (Consumer Research)*. Hrsg. Forschungsgruppe Konsum und Verhalten, 107–117, Munchen: Verlag Franz Vahlen.
- Buck, R., E., Anderson, A., Chaudhuri, and I., Ray. 2004. Emotion and reason in persuasion: Applying the ARI Model and the CASC Scale. *Journal of Business Research. Marketing Communications and Consumer Behavior* 57: 647–656.
- Buck, R., A., Chaudhuri, M., Georgson, and S., Kowta. 1995. Conceptualizing and operationalizing affect, reason, and involvement in persuasion: The ARI model and the CASC scale. *Advances in Consumer Research* 22: 440–447.
- Carter, C. S., I. I. Lederhender, and B. Kirkpatrick (eds.). 1997. *The Integrative Neurobiology of Affiliation. Annals of the New York Academy of Sciences. Vol. 807*. New York: The New York Academy of Sciences.
- Chaiken, S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39: 752–766.
- Chaiken, S. 1987. The heuristic model of persuasion. In *Social Influence: The Ontario Symposium*. M. P. Zanna, J. M. Olson, and C. P. Herman, eds., 3–39, Hillsdale, NJ: Erlbaum.
- Chaiken, S., and A. H., Eagley. 1983. Communication modality as a determinant of persuasion: The role of communicator salience. *Journal of Personality and Social Psychology* 45: 241–256.

- Chaudhuri, A., and R., Buck. 1993. The relationship of advertising variables to analytic and synthetic cognitions. In *Marketing Theory and Applications*. R. Varadarajan, and B. Jaworski, eds., 193–198, Chicago, IL: American Marketing Association.
- Cook, S., L., Peterson, and D., DiLillo. 1999. Fear and exhilaration is response to risk: An extension of a model of injury risk in a real-world context. *Behavior Therapy* 30: 5–15.
- Cox, E. P., M. S., Wogalter, and S. L., Stokes. 1997. Do product warnings increase safe behavior? A meta-analysis. *Journal of Public Policy and Marketing* 16: 195–204.
- Deppa, S. W., and B. J., Martin. 1997. Human factors behind the improved ANSI Z535.3 label standard for safety symbols. *Proceedings of the Human Factors and Ergonomics Society*. 41st Annual meeting.
- deTurck, M. A., G. M., Goldhaber, and G. M., Richetto. 1993. Familiarity and awareness: Effects of conscious and nonconscious safety information. *Journal of Products Liability* 14: 341–350.
- deTurck, M. A., G. M., Goldhaber, and G. M., Richetto. 1995. Effectiveness of alcohol beverage warning labels: Effects of consumer information processing objectives and color of signal word. *Journal of Products and Toxics Liability* 17: 187–195.
- Eckerman, I. 2004. *The Bhopal Saga – Causes and Consequences of the World's Largest Industrial Disaster*. India: Universities Press.
- Fischer, P. M., J. W., Richards Jr., E. J., Berman, and D. M., Krugman. 1989. Recall and eye tracking study of adolescents viewing tobacco advertisements. *Journal of the American Medical Association* 261(1): 84–89.
- Gabrielsen, M. A., and M., Spivey. 1990. *Diving Injuries: The Etiology of 486 Case Studies with Recommendations for Needed Action*. Fort Lauderdale, FL: Nova University Press.
- Givel, M. 2007. A comparison of the impact of US and Canadian cigarette pack warning label requirements on tobacco industry profitability and the public health. *Health Policy* 83: 343–352.
- Hammond, D., G. T., Fong, R., Borland, K. M., Cummings, A., McNeill, and P., Driezen. 2007. Text and graphic warnings on cigarette packages: Findings from the International Tobacco Control Four Country Study. *American Journal of Preventive Medicine* 32: 202–209.
- Hastings, G., and L., MacFadyen. 2002. The limitations of fear messages. *Tobacco Control* 11: 73–75.
- Kosfeld, M., M., Heinrichs, P. J., Zak, U., Fischbacher, and E., Fehr. 2005. Oxytocin increases trust in humans. *Nature* 435: 673–676.
- Lazarus, R. 1991. Progress on a cognitive-motivational-relational theory of emotion. *The American Journal of Psychology* 46: 819–834.
- LeDoux, J. E. 1994. Memory versus emotional memory in the brain. In *The Nature of Emotion: Fundamental Questions*. P. Ekman, and R. J. Davidson, eds., 311–312, New York: Oxford University Press.
- LeDoux, J. E. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.
- MacKinnon, D. P., and L., Nohre. 2006. Alcohol and tobacco warnings. In *Handbook of Warnings*. M. S. Wogalter, ed., 669–685, Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- MacLean, P. D. 1993. The cerebral evolution of emotion. In *Handbook of Emotions*. M. Lewis, and J. Haviland, eds., New York: Guilford Press.
- Martin, B. J., and S. W., Deppa. 1997. Human factors in the revised ANSI Z535.4 standard for safety labels. *Proceedings of the Human Factors and Ergonomics Society*. 41st Annual meeting: 821–825.
- McGrath, J. M., and C. W., Downs. 1992. The effectiveness of on-product warning labels: A communication perspective. *For the Defense* 1992: 19–24.
- Petty, R. E., and J. T., Cacioppo. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer Verlag.
- Panksepp, J. 1993. Neurochemical control of moods and emotions: Amino acids to neuropeptides. In *Handbook of Emotions*. M. Lewis, and J. Haviland, eds., New York: Guilford Press.

- Panksepp, J. 1994. A proper distinction between cognitive and affective process is essential for neuroscientific progress. In *The Nature of Emotion: Fundamental Questions*. P. Ekman, and R. J. Davidson, eds., 224–226, New York: Oxford University Press.
- Peters, E., D., Romer, P., Slovic, K. H., Jamieson, L., Wharfield, C. K., Mertz, and S. M., Carpenter. 2007. The impact and acceptability of Canadian-style cigarette warning labels among US smokers and nonsmokers. *Nicotine & Tobacco Research* 9: 473–481.
- Piaget, J. 1971. Piaget's theory. In *Handbook of Child Development, Vol. I*, P. Mussen, ed., New York: Wiley.
- Ramakant, B. 2008. Pictorial warnings on tobacco products most likely postponed 7th time, India. *Medical News Today*. Online at <http://www.medicalnewstoday.com/articles/130966.php> (Accessed 2008/11/27)
- Sherif, M. 1965. Formation of social norms. In *Basic Studies in Social Psychology*. H. Proshansky, and B. Seidenberg, eds., 461–471, New York: Holt, Rinehart & Winston.
- Tucker, D. M. 1981. Lateral brain function, emotion, and conceptualization. *Psychological Bulletin* 89: 19–46.
- White, R. W. 1959. Motivation reconsidered: The concept of competence. *Psychological Review* 66: 297–333.
- Wogalter, M. S., and K. R., Laugherty. 1996. WARNING! Sign and label effectiveness. *Current Directions in Psychological Science* 5: 33–37.
- Wogalter, M. S., S. S., Fontenelle, D. R., Desaulniers et al. . 1987. Effectiveness of warnings. *Human Factors* 29: 599–612.
- Zuckerman, A., and S., Chaiken. 1998. A heuristic-systematic processing analysis of the effectiveness of product warning labels. *Psychology and Marketing* 15: 621–642.
- Zuckerman, M. 2007. *Sensation Seeking and Risky Behavior*. Washington, DC: American Psychological Association.

Emotions as Aids and Obstacles in Thinking About Risky Technologies

Dylan Evans

Developments in technology have prompted ethical concerns for as long as recorded history. Writing itself is denounced by Plato in the *Phaedrus*, and other technological developments since then that have attracted moral censure include the mechanical clock, the crossbow, printing, the steam engine, vaccinations, and nuclear power, to name only the most notorious examples. It is as if the extent of man's curiosity and genius for invention were equalled only by his apparent discomfort with these faculties. This discomfort is encoded in many ancient myths, from the Hebrew story of the expulsion from the Garden of Eden, to the Greek tale of Prometheus and the Mayan legend of the rebellion of the tools.

Although contemporary developments such as genetic engineering, nanotechnology, and the use of stem cells in medical research are new, there is nothing new, therefore, about the aversion that many people today feel towards new technology. Indeed, it is all depressingly familiar.

What is new is a certain willingness by some scholars to endow this aversion with some normative weight. Traditionally, philosophers (in the Western tradition at least) have regarded emotional reactions as inimical to rational appraisal. In his famous metaphor of the chariot, Plato portrayed the passions as horses and reason as the charioteer. The message is clear; the passions may provide motive power, but it is up to the charioteer to steer them in the right direction. Immanuel Kant too argued that moral decisions should be a matter for pure reason, excluding all "pathological" emotional considerations. In standard accounts of the history of Western philosophy, Kant's views are usually contrasted with those of David Hume, who argued that approbation or blame "cannot be the work of the judgement, but of the heart; and is not a speculative proposition or affirmation, but an active feeling or sentiment" (Hume 1777 Appendix 1). However, it is worth noting that Kant and Hume agree on a fundamental idea; namely, that moral judgements are irrational to the extent that they are determined by emotional considerations. Kant believes that moral judgements can and should be made without emotional involvement, and

D. Evans (✉)

Behavioural Science, School of Medicine, University College Cork, Cork, Ireland
e-mail: evansd66@googlemail.com

are rendered irrational to the extent that they are contaminated by emotion. Hume believes that moral judgements are always determined by emotional considerations, and concludes that they are therefore irrational (or at least arational). Neither Kant nor Hume attribute any normative weight to our emotional reactions.

Recently, however, some thinkers have proposed an alternative view, according to which emotions can be a normative guide in making moral judgments. Perhaps the best known proponent of this view is Leon Kass, who argues that feelings of disgust may be the manifestation of a kind of moral “wisdom” (Kass 1997). Kass is certainly not alone, however, and nor is disgust the only emotion that these attempts to endow emotions with normative weight have focused on. Roeser, for example, agrees with Kass that emotions can be a normative guide in making moral judgements, but her focus is on sympathy, empathy, fear and indignation (Roeser 2006b).

Critics of Kass have rightly pointed out that human history is littered with examples of things that were once considered disgusting but which we now recognise were inappropriate objects of revulsion. Homosexuality, working women, and other races were all considered disgusting by very large numbers of people, and sometimes whole societies. Yet few would say today that those feelings were appropriate. As John Harris points out, “we ought to have a rational caution about following the yuk factor because we know it has led us not only in the wrong direction but in a thoroughly corrupt direction” (cited in Ahuja 2007). History teaches us that we cannot rely on the emotion of disgust to provide our moral compass. Like other emotions, disgust can be educated, but it can also have dubious causes.

The same arguments could be made, of course, against claims for the moral significance of other emotions besides disgust such as Roeser’s claims for the moral significance of sympathy. More important than any claims about the moral significance of particular emotions such as disgust or fear, though, is the logically prior claim that emotions of whatever kind can carry normative weight. To my mind, this is Kass’s most fundamental error; the argument about disgust is important only as a special case of the more general claim.

Prima facie, it would seem that Kass and the other thinkers who share his views on the normative weight of emotions are simply making an elementary philosophical blunder by failing to observe the is-ought distinction. If one starts with the premise that research involving embryonic stem cells is disgusting, and concludes (after any number of intermediate steps) that one ought not to engage in such research, then it is clear that at some point in the argument one has made an invalid inference unless one of those intermediate steps is a premise to the effect that one ought not to do disgusting things. It is then clear that the moral weight of the argument depends on this crucial moral claim, and not on the empirical facts.

However, the proponents of the moral emotion view (as I shall call it here) would presumably reject such a criticism on the grounds that it is too simplistic. Roeser, for example, claims to base her views on “recent developments in neurobiology, psychology and the philosophy of emotions”, which, she thinks, show that “emotions and rationality are not mutually exclusive, but rather, in order to be practically rational, we need to have emotions” (Roeser 2007). Roeser takes these empirical findings in psychology to provide some support for her specific version of the

philosophical position known as ethical intuitionism – the thesis that we sometimes have intuitive awareness of value, or intuitive knowledge of evaluative facts, which forms the foundation of our ethical knowledge. According to Roeser, ethical intuitions are paradigmatically cognitive moral emotions with which we perceive objective moral truths (Roeser 2006a).

Moral realism is the critical premise on which all of Roeser's claims about the normative status of emotions depend. To refute these claims decisively, then, it would be necessary to show that moral realism is false. Limitations of space make it impossible, however, to rehearse the well-known and well-established arguments against this notion here. For the purpose of this article, I will limit myself to dealing only with what Roeser herself calls "the main argument for moral realism" (Roeser 2006b, p. 692). This argument begins by assuming that if there were no moral truths, there would not be an objective standard against which to evaluate a situation. It then appeals to our moral intuitions, which tell us clearly that certain moral practices are wrong, and by *modus tollens* infers that there must be moral truths. This is a valid argument, but the conclusion is only true if one accepts that our moral intuitions are good guides to the truth. Yet this is exactly what the argument purports to show, so the reasoning is circular. "This might sound like wishful thinking or circular reasoning," admits Roeser, but then adds; "it is rather to be understood as 'inference' to the best explanation" (Roeser 2006b, p. 692). This is disingenuous; no amount of denial will obscure the blatant circularity.

Nor does the occasional reference to "cognitive" theories of emotion provide any support for any species of moral realism. I suppose Roeser is right to claim that "cognitive theories of emotions allow for the idea that emotions are basic perceptions of moral reality" (Roeser 2006b, p. 692), but the mere fact that cognitive theories of emotion might be logically consistent with the thesis of moral realism does not provide any grounds for thinking that it is true. Some cognitive theories of emotion hold that emotions are judgements of value (Nussbaum 2001), but the sense in which the term "value" is used here is not a moral or ethical one. Rather, what "value" means in this context is the relation that some event or fact has to an organism's desires or intentions. Something has value in this sense if and only if it is either a potential aid or a potential obstacle to the achievement of one's desires or intentions, irrespective of any moral or ethical matters. It is simply a category mistake, therefore, to think that cognitive theories of emotion provide any support for any species of moral realism.

It is likewise a mistake to think that certain contemporary views on the role that emotions play in practical reasoning have any bearing on the question of whether or not emotions have normative weight. The view that humans need emotions in order to be practically rational has become increasingly popular in the past decade (eg. de Sousa 1987; Evans 2002). But, like the cognitive theories of emotion with which this view is often closely associated, it is entirely a matter of empirical psychology, and has no necessary link with any species of moral realism.

In what follows, then, I simply assume that values, norms and ethics are all subjective phenomena, in the sense that we may have opinions about them, but there are no facts of matter. This does not imply, of course, that no practice can be morally

better or worse than another. It simply means that statements about the relative moral value of different practice must always be relativised to a given person or community. Democracy may be morally better than dictatorship for this person, or for that community, but never per se.

According to Roeser, her philosophical framework “is meant as a ‘third way’ between Kant and Hume” (personal communication). But like many “third ways”, this is really no more than incoherence masquerading as complexity. The truth does not always lie half-way between two opposing views. Sometimes, the dichotomy exhausts the space of logical possibilities. In such cases, to reject the dichotomy as being “too simplistic” is intellectually dishonest. It muddies the water and prevents clear debate.

Until Roeser and the other proponents of the moral emotion view provide a clearly articulated explication of their much-vaunted “third way”, then, we must treat their claims as mere hand-waving. Neither they nor anyone else has yet provided sufficient reasons to question the widely held view that emotions provide no evidence at all for or against any moral or ethical claim.

Does this mean that emotions convey no ethical or moral information? Certainly not. Emotional reactions often (but not always) convey information about the ethical and moral beliefs of the person exhibiting the reaction. If a person reacts with anger when she reads about a businessman who retires with a fat pension after almost bankrupting his company, I can reasonably infer that among her moral beliefs is one that places a high value on accountability and fairness.

Although this idea is hardly new, it is still underdeveloped. When combined with recent developments in psychological ethics, however, such as Jonathan Haidt’s “moral foundations theory”, it gives rise to some interesting consequences.

Haidt argues that there are five psychological systems that provide the foundations for the world’s many moralities. Each system is specialised for detecting and reacting emotionally to distinct issues: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity. When the harm/care system is triggered, the emotions of fear and compassion may be activated. The fairness/reciprocity system evokes primarily the emotions of anger, gratitude and guilt. The ingroup/loyalty system involves strong social emotions related to recognizing, trusting, and cooperating with members of one’s co-residing ingroup, while being wary and distrustful of members of other groups. Emotions of pride, shame, awe and admiration, are manifestations of the authority/respect system. Finally, activation of the purity/sanctity system is associated most strongly with the emotion of disgust:

Disgust appears to function as a guardian of the body in all cultures, responding to elicitors that are biologically or culturally linked to disease transmission (feces, vomit, rotting corpses, and animals whose habits associate them with such vectors). However, in most human societies disgust has become a social emotion as well, attached at a minimum to those whose appearance (deformity, obesity, or diseased state), or occupation (the lowest castes in caste-based societies are usually involved in disposing of excrement or corpses) makes people feel queasy. In many cultures, disgust goes beyond such contaminant-related issues and supports a set of virtues and vices linked to bodily activities in general, and religious activities in particular. Those who seem ruled by carnal passions (lust, gluttony, greed, and anger) are seen as debased, impure, and less than human, while those who live so that

the soul is in charge of the body (chaste, spiritually minded, pious) are seen as elevated and sanctified. (Haidt and Graham 2007, p. 116)

Haidt's theory allows us to make much more systematic inferences about the information that emotional reactions often convey about the ethical and moral beliefs of the person exhibiting the reaction. For example, if someone appeals to the emotion of fear when expounding on their moral opposition to GM crops, we can infer that the risks they associate with this technology are largely to do with the possible harm that this technology could do (by, for example, damaging the digestive system of those who consume them). Alternatively, if the emotion of anger plays a larger role in someone's opposition to GM crops, we might infer that the risks they associate with this technology have more to do with possible injustice (such as increasing the profits of large corporations at the expense of small farmers). Or, again, if it is the emotion of disgust that seems to motivate the opponent of GM crops, it may be that the risks that weigh most heavily on their mind are spiritual or theological ones (such as "tampering with God's creation").

Haidt has also argued that political liberals tend to base their moral intuitions primarily upon just two systems (the harm/care and fairness/reciprocity systems), while political conservatives generally rely upon all five systems. Liberals therefore often misunderstand the moral motivations of conservatives, explaining them as a product of various non-moral processes such as system justification or social dominance orientation. The fact that bioconservatives like Kass see wisdom in the emotion of disgust is clearly in line with Haidt's claim that the values of purity and sanctity tend to play an especially important role in the moral beliefs of political conservatives. Similarly, the fact that liberals like Harris disparage the appeal to this emotion is also in line with Haidt's view that purity and sanctity do not even figure as concepts in liberal moral systems.

Haidt's thesis is not necessarily disproven by the recent appropriation of disgust by liberal thinkers. Dan Kahan, for example, has argued that even a liberal society needs to build law on the basis of disgust and attempts "to redeem disgust in the eyes of those who value equality, solidarity, and other progressive values" (Kahan 2000). Liberals should not, he argues, cede the "powerful rhetorical capital of that sentiment to political reactionaries" just because prominent defenders of disgust have often used it to defend conservative ideas. While this may seem an interesting tactical manoeuvre, if Haidt is right about the deeper psychological foundations of moral discourse, it is not likely to win much support among liberals. Time will tell.

Haidt's analysis is valuable here, not just because of his theses about specific emotions such as disgust, but also for the more general light that it throws on the debate about the role of emotions in moral reasoning. Perhaps the debate between Harris and Kass is not about the importance of emotion per se in moral reasoning, but about the relative value of particular emotions in moral reasoning. If this is the case, then it might be more perspicuous to view the debate between Harris and Kass, not as simply a rerun of the Kant/Hume debate, with Harris playing the role of Kant and Kass the role of Hume, but rather as a debate between different species of Humean ethics. If this is true, we would expect Harris and Kass to agree on the

importance and relevance of emotions like compassion and pity to moral debate, since both liberals and conservatives base their moral intuitions on the harm/care system with which such emotions are associated. A true Kantian, of course, would take these emotions to be just as irrelevant to moral reasoning as the emotion of disgust.

Even a Kantian can, however, find something of value in this analysis. The fact that a person's emotional reaction can be used to infer their implicit moral values does not, of course, imply that emotions carry any normative weight. The Kantian is nevertheless perfectly entitled to avail himself of such emotional evidence to help tease out the moral values which are at stake in the argument. Once emotions have been used in this way, the argument can proceed in an entirely unemotional way.

In the case of arguments about risky technologies, the Kantian can use the evidence provided by emotional reactions to help clarify what exactly the risks are that a person associates with a given technological development. When this has been established, however, the likelihood of those risks will be assessed by rational means alone – that is, by statistical evidence, without reference to emotion-laden perceptions. For example, suppose that my reaction to some new development in biotechnology is fear – fear that the acceptance of this vital new technology may be hampered by misleading propaganda put about by environmentalists. That would suggest that the risks that matter most to me are risks of possible harm – in this case, the harm done to humanity by depriving people of a means for improving quality of life – rather than the risk of injustice or some imaginary “theological risk”. That, in my view, is where the “moral” issues end. What remains is for me to gather empirical data about the likelihood of the risks I care about. This is a purely statistical matter.

The findings that have accumulated over four decades of research in the heuristics and biases programme must remain the key reference point here. These findings show conclusively that emotions almost always tend to reduce the rationality of decisions regarding the moral acceptability of technological risks by causing us to pay more attention to potential harms or potential benefits than is warranted by the evidence. Sometimes, enthusiasm can lead proponents to pay too much attention to the benefits of a technology and not to pay enough attention to risks. More often, however, it is the other way round, with the risks getting too much attention and the benefits being downplayed. The prevalence of this “luddite bias” may have some evolutionary basis; many emotional subsystems in the brain seem to be biased in the direction of perceiving threats at the expense of missing benefits, and overall there are many more negative emotions than positive ones. Thus people tend to be better at imagining the potential harm of new technologies than imagining the benefits.

As Cass Sunstein has pointed out in *Laws of Fear*, a truly rational analysis will always balance the risks of developing a given technology against the risks of not developing that technology (Sunstein 2005). The luddite bias is therefore an obstacle to rational analysis. One may attempt to overcome this obstacle by systematic debiasing methods, such as forcing oneself to list as many potential benefits as potential harms when considering a new technology. Given the powerful emotional nature of the luddite bias, however, intellectual corrective procedures may not be enough

to counteract it, and it may therefore be necessary to employ emotional debiasing techniques too. For example, one might attempt to elicit the corresponding positive emotion for each negative emotion. When considering the possibility that GM foods might be toxic, for example, we should also consider the possibility that they might help avert starvation in developing countries, and we should try to elicit the emotion of compassion for the millions of people who might be helped in this way. Alternatively, if we are carried away by enthusiasm for a particular technological development, we might try to elicit a reasonable degree of fear for the potential risks.

This process is not, of course, a substitute for the rational assessment of the likelihood of the potential harms and benefits, but merely attempts to make sure that the emotional input into the decision-making process is fair and balanced and so less likely to distort the unbiased gathering of relevant information.

Conclusion

I have outlined a way in which emotions may play a role in assessing the moral acceptability of the risks associated with new technologies which does not impair the rationality of such assessments. Even a Kantian could claim that emotions could enhance the rationality of such assessments. This underlines the importance of spelling out precisely the nature of claims about the “rationality of emotion”, which can cover a multitude of sins. All a Kantian would mean by such a phrase is that emotions can help to clarify what exactly the risks are that a person associates with a given technological development. Their role is purely to provide empirical evidence concerning the implicit values of a given person.

Acknowledgments I am grateful to Sabine Roeser for her comments on several previous versions of this manuscript.

References

- Ahuja, A. (2007) *Enhancing the Species*. The Times. 10 October 2007 Accessed on 22 April 2009, from http://women.timesonline.co.uk/tol/life_and_style/women/families/article2622232.ece
- de Sousa, R. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press.
- Evans, D. 2002. The search hypothesis of emotion. *The British Journal for the Philosophy of Science* 53(4): 497–509.
- Haidt, J., and J., Graham. 2007. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research* 20: 98–116.
- Hume, D. (1777). *An Enquiry into the Principles of Morals*. Adelaide, eBooks@Adelaide. <http://ebooks.adelaide.edu.au/h/hume/david/h92pm/complete.html> Accessed on 23 April 2009.
- Kahan, D. M. 2000. The progressive appropriation of disgust. In *The Passions of Law*. S. Bandes, ed., New York: New York University Press.
- Kass, L. R. 1997. The wisdom of repugnance. *The New Republic* 216: 17.
- Nussbaum, M. 2001. *Upheavals of Thought: The Intelligence of Emotions*. New York: Routledge.
- Roeser, S. 2006a. A particularist epistemology: “Affectual intuitionism”. *Acta Analytica* 21(1): 33–44.

- Roeser, S. 2006b. The role of emotions in judging the moral acceptability of risks. *Safety Science* 44: 689–700.
- Roeser, S. (2007). *Conference 'Moral Emotions about Risky Technologies'*. Retrieved 18 April 2009, 2009, from http://www.ethicsandtechnology.eu/news/comments/moral_emotions_about_risky_technologies/
- Sunstein, C. R. 2005. *Laws of Fear: Beyond the Precautionary Principle*. Cambridge: Cambridge University Press.

Part II
Emotions and Virtues in Risk Assessment

Risk Assessment as Virtue

Sabine Döring and Fritz Feger

1 Introduction

In the following, we shall present a critique of decision theory as a normative account of decision making under risk. We claim that decision theory has to be supplemented by virtue. To some of you, speaking of “virtue” might sound old-fashioned or even humorous. But we use it as a technical term that refers to a person’s capability to assess risk appropriately in an immediate, non-inferential way, rather than intellectually calculating and thus inferring risk. Virtuous risk assessment manifests itself both in its possessor’s sensibility towards risk and in his being motivated to act accordingly. As such, it is not simply a skill, but an expression of practical wisdom ($\varphi\rho\rho\nu\eta\sigma\iota\varsigma$) which equips humans for managing the complexity of real life, thereby aiming to enhance the quality of life. In this very broad sense, virtuous risk assessment is ethically salient and thus an ethical (rather than a rational) virtue. But it need not be morally salient, i. e. concern the question of “what we owe to each other”, to borrow Scanlon’s (1998) instructive characterisation of this much narrower domain. To establish that decision making under risk requires virtue, we shall, first, show that decision theory is unable to resolve the well-known St. Petersburg paradox. The St. Petersburg game poses a long standing problem to decision theory because it has infinite expected value and yet seems to be worth much less. As we shall argue in a second step, the systematic deviation from apparently rational choice can be justified as an instance of virtue. This means that, by contrast with the inferential method of risk assessment characteristic of decision theory, virtuous risk assessment requires having appropriate emotions. An emotion, it is assumed, is a perception of value. Our third aim, then, is to make plausible that, contrary to common belief, decision situations can be found in real life which entertain the structure of the St. Petersburg game, or rather of the inverted St. Petersburg game, which we shall introduce. We maintain that in many cases the

S. Döring (✉)
Philosophisches Seminar, Universität Tübingen, Tübingen, Germany
e-mail: mail@sabinedoering.de

assessment of risky technologies is such that a decision has to be made about possible but extremely unlikely outcomes with an “infinite” negative value (so to speak). In these cases, virtue is needed to avoid an inappropriate assessment of these options by a decision theory that is expected to do too much.

2 The St. Petersburg Paradox

Consider the following game and ask yourself how much you would be willing to invest to enter it. The game, known as the St. Petersburg game, is played by flipping a fair coin until it comes up tails. When this happens the game ends. Let n denote the total number of flips until the coin comes up tails. Your prize is determined by n , which equals $\$2^n$. Thus, if the coin comes up tails the first time, the prize is $\$2^1 = \2 , and the game ends. If the coin comes up heads the first time, it is flipped again. If it comes up tails the second time, the prize is $\$2^2 = \4 , and the game ends. If the coin comes up heads the second time, it is flipped again. And so on. Since infinitely many runs of heads are possible until the coin comes up tails the first time, the number of possible outcomes of the game is infinite. Table 1 below lists the outcomes for $n = 1 \dots 10$, their probabilities, and the resulting expected payoff for each outcome:

Table 1 The St. Petersburg game

n	P(n)	Prize (\$)	Expected payoff (\$)	Cumulative expected payoff (\$)
1	1/2	2	1	1
2	1/4	4	1	2
3	1/8	8	1	3
4	1/16	16	1	4
5	1/32	32	1	5
6	1/64	64	1	6
7	1/128	128	1	7
8	1/256	256	1	8
9	1/512	512	1	9
10	1/1024	1024	1	10

As can be seen from the table, the point of the St. Petersburg game is that the prize equals the reciprocal value of the respective probability such that the expected payoff for each possible outcome is the same. The higher the value of n , the lower its probability. But the decreasing probability is exactly compensated by an increasing prize, which is reflected by the constant expected payoff. The expected value of the whole game equals the sum of the expected payoffs of all the outcomes. As there are infinitely many possible outcomes, the constant expected payoff of each outcome adds up to an infinitely high expected value for the whole game:

$$E = \sum_{n=1}^{\infty} \left(\frac{1}{2^n} \cdot 2^n \right) = \sum_{n=1}^{\infty} 1 = \infty$$

According to the theory of rational choice, or decision theory, in its most basic and simple version, a rational gambler enters a game iff the entrance fee is smaller than the expected value of the game. In the St. Petersburg game, any finite entrance fee is smaller than the expected value of the game. Thus, the rational gambler plays no matter how high the stakes.

The question is whether you would be willing to pay any finite entrance fee, such as your annual income, say. Probably not. In most cases, you will win only a couple of dollars, and half of the games are worth no more than \$2. In his article *Strange Expectations* Hacking (1980) suggested that “few of us would pay even \$25 to enter such a game”. Most commentators agree. There seems to be a common “intuition” that one should not even spend \$25 to enter the St. Petersburg game. If this is correct, decision theory is challenged. By generating a random variable with an infinite expected outcome which seems to be worth much less, the St. Petersburg game poses a problem. Therefore it is also called the “St. Petersburg paradox”.

The challenge to decision theory is twofold. Decision theory is expected to explain behaviour, and much of its interest depends on its explanatory power. This expectation is disappointed by the St. Petersburg game, because here theory predicts a choice people do not make. A deviation from apparently rational choice is observed, which cannot be neglected as randomly distributed but occurs systematically. Thus the St. Petersburg paradox shows, first, that decision theory is descriptively inadequate.

Secondly, the normative adequacy of decision theory is challenged. Decision theory is an explanatory model which includes the rationality of the agent among its assumptions. It is about the choices of an ideally rational agent, not about the various ways in which people may happen to make irrational decisions. Accordingly, it does not only describe or predict what a rational agent does. Decision theory involves a claim about what an agent ought to do in order to be rational. However, in the St. Petersburg game, a choice is qualified as rational which rational agents ought not to make. It seems foolish to pay more than a modest sum of money for the game. Thus the St. Petersburg paradox shows, secondly, that decision theory is normatively inadequate.

In the following, our concern will be with normative adequacy. Decision theory will be treated as a method for providing normative reasons for action, namely, the method of using expected payoffs to evaluate options. In accord with this, we shall take the St. Petersburg paradox as a methodological problem, which cannot be resolved simply by questioning its “realism”. Arguing, for example, that there is no limitless supply of money for expected payoffs, or that it is physically impossible to flip a coin infinitely many times, or that, given the constraints of real life, no one will ever offer us the game, will not do. Whether or not anyone will ever be confronted with the St. Petersburg game in real life: should this happen, decision theory is required to provide a satisfactory answer to the question of what a rational

agent ought to do. Taken as a logical challenge to the method of rational decision making, the St. Petersburg paradox is “impeccable”, as Resnick (1987) puts it in his *Introduction to Decision Theory*.

Although our focus will be on the normative dimension of decision theory, this is not to ignore factual deviations from what theory prescribes. One might think that, from a normative point of view, factual deviations can be ignored: perhaps it is rational to pay for the St. Petersburg game according to expected payoff, even though in reality we fail to use this criterion of rational choice. In our view, this line of argument amounts to dogmatic normativism, if, as in the St. Petersburg paradox, factual deviations are systematic and express a commonly shared intuition. Systematic factual deviations from rational choice, also referred to as “puzzles” or “anomalies”, need to be explained. The required explanation may preserve decision theory by identifying a systematically misleading causal mechanism, which is held responsible for the anomaly. (This is similar to, e. g., explaining the well-known Müller-Lyer illusion by the size constancy mechanism of our visual system.) If this is not possible or implausible, decision theory is at stake: either it must be revised, or it has to be substituted or at least supplemented by an alternative method of rational choice. Figure 1 below illustrates the available ways to analyse factual deviations from what decision theory predicts.

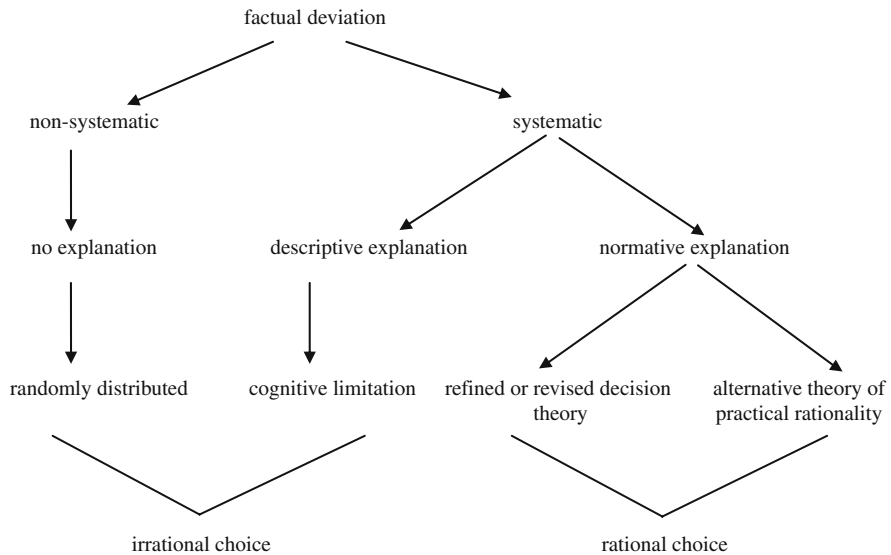


Fig. 1 Explications of factual deviations from what decision theory predicts

We shall show now that our intuitions about the St. Petersburg game cannot be captured by decision theory, and yet they do not betray us.

3 Decision-Theoretic Attempts to Resolve the St. Petersburg Paradox

All attempts to meet the methodological challenge of the St. Petersburg paradox share one feature: a decision theory based on “straight” expected value is dismissed as naive. Instead, the St. Petersburg problem is tackled by formulating models of how human agents evaluate payoffs or probabilities. The first and most influential proposal along this line of argument is Bernoulli’s (1738/1954) now famous introduction of the concept of expected utility. Rather than relying on expected payoff as such, Bernoulli says, we should use the utility generated by expected payoff to evaluate options. He introduces the law of the diminishing marginal utility of money, which states that the additional utility generated by an additional money unit decreases. One additional dollar means more to me when I am poor than when I am rich. This is a hypothesis regarding the shape of the utility function. The utility function is claimed to be concave from below. If this is so, the constant expected payoff of each outcome yields a decreasing expected utility. Marginal utility remains greater than zero (goods are good), but, if hypothesised to decrease, necessarily approaches zero for amounts of money approaching infinity. Therefore, the expected utility of the game approaches a finite value.

Bernoulli’s law of the diminishing marginal utility of money is now an inherent part of the theory of rational choice. However, refining expected value in terms of expected utility does not suffice to resolve the St. Petersburg paradox. As Menger (1934) points out, it is always possible to compensate payoff so that expected utility remains infinite. We then get the “Super St. Petersburg game”, which has constant marginal utility, rather than constant marginal payoff, and is thus immune against any attempt to resolve the paradox by modifying the shape of the utility function. Menger’s own solution is to postulate an upper bound on utility, which he thinks is the only way that the paradox can be resolved. Some have seconded his view (cf. Hardin 1982; Jeffrey 1983; Gustason 1994). Pace Menger, we find this classical treatment of the St. Petersburg paradox ad hoc and not to the point. The idea of an upper limit to utility amounts to explaining away a logical paradox by rejecting one of its factual assumptions. Anyone who understands the problem appreciates that, apart from his own intuitive choice, there is the choice of an ideally rational agent with unbounded utility who is prepared to pay any finite sum. Furthermore, it is empirically questionable to assume that, on top of diminishing marginal utility, there should be some amount of utility which is so high that no additional utility is possible. The more, the better, we should say (cf. also Martin 2004).

A more appealing strategy to resolve the St. Petersburg paradox refers to our attitude towards risk (cf., e. g., Friedman and Savage 1948; Arrow 1964; Pratt 1964; Weirich 1984). Unfortunately, the most common account of risk aversion, shown in Fig. 2 below, is again subject to Menger’s objection. Since in the expected utility framework diminishing marginal utility is the sole explanation for risk averse behaviour, prizes can again be adjusted so as to render the expected utility of every possible outcome constant (cf. Rabin 2000). The paradox stays alive and well.

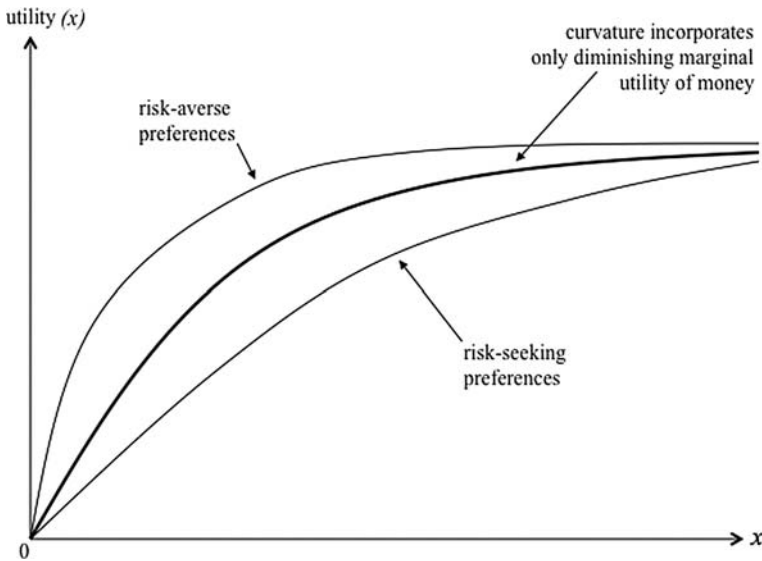


Fig. 2 Attitude towards risk in terms of the curvature of the utility function

Weirich (1984) treads a different path by adopting the so-called mean-risk method (cf. Markowitz 1959). According to Weirich, although probability is included in the calculation of expected utility, the latter does not account for the agent's attitude towards risk. Weirich believes that risk causes “negative utility”, where risk consists in low probability as such. The negative utility of risk is subtracted from the positive utility of expected payoff. Applying the mean-risk method to the St. Petersburg game in this way, Weirich says, yields a net expected utility for the whole game which meets the common intuition.

The underlying assumption of his account is that, in the St. Petersburg game, the probabilities of the prizes make the value of the game finite no matter what values the prizes have (cf. Weirich 1984, 194 f.). While we share this assumption, we find its implementation implausible. To be sure, we hold a high opinion of the rational agent in expecting him to get the point of the St. Petersburg paradox. But Weirich seems to be asking too much. Within his model, to subtlety a calibration of parameters is required to guarantee the desired result. To subtract a disutility of risk of round about minus infinity from an expected utility of round about plus infinity so as to land in the tiny range between \$2 and \$25 sounds like magic. Against this background, it is questionable whether the criterion of rational choice proposed by Weirich in order to resolve Menger's generalisation of the St. Petersburg paradox is of any relevance in other decision situations involving risk. Kahneman's and Tversky's (1979, 1992) empirical findings on decision making under risk point to a different direction. They observe that we tend to underweight “average” events, while we attach too much importance to extreme, but relatively unlikely events.

This contradicts Weirich’s claim that low probability makes us reluctant to consider an option.

It must however be noted that Kahneman and Tversky do not deal with the St. Petersburg paradox. Their “cumulative prospect theory” (and its predecessor “prospect theory”) is designed to explain a number of other anomalies. In doing so, Kahneman and Tversky rely on the classical strategy of modeling risk in terms of evaluating expected payoff. By contrast with Weirich, risk aversion is not analysed as an evaluation of probability as such. The main departure from expected utility theory is in another respect: changes of wealth, rather than absolute levels of wealth, are regarded as carriers of value. This is to capture the empirical fact that the value which we attach to a possible outcome depends on a certain reference point (often the status quo), rather than on the absolute status, a phenomenon which is called the “framing effect”. The modification is significant because it makes possible to account for “loss aversion”, i. e., for the fact that responses to losses are much more intense than responses to corresponding gains. The rage over the lost penny is bigger than the joy at the penny found in the street, as we may put it. We then arrive at a value function as depicted in Fig. 3 that is concave for gains (implying risk aversion), convex for losses (implying risk tolerance), and steeper for losses than for gains (implying loss aversion).

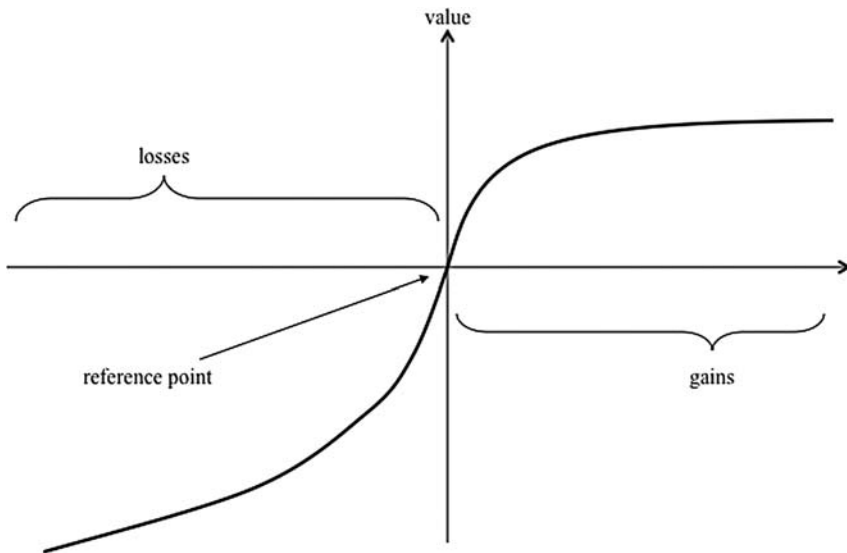


Fig. 3 Value function according to Kahnemann and Tversky

We have already pointed out that the classical strategy of modeling risk in terms of the shape of the utility function fails to resolve the St. Petersburg paradox because it cannot escape Menger’s objection. In Kahneman’s and Tversky’s model this strategy seems to do a good job in explaining other anomalies of decision making under risk. But, applied to the St. Petersburg paradox, cumulative prospect theory clearly

fails: the paradox is not resolved, but even intensified. Presuming that we tend to overweight extreme, but unlikely outcomes, our willingness to pay any finite sum for the game should even be greater than predicted by standard decision theory. Observed probability weights and value function parameters yield infinite subjective values even for games with finite expected value (cf. Rieger and Wang 2004; Blavatsky 2005).

Kahneman and Tversky hesitate to draw any normative conclusions from their empirical findings (cf., however, Tversky and Kahneman 1992). On our view, the normative adequacy of standard decision theory is challenged by the fact that the St. Petersburg paradox is intensified, rather than resolved, by the most up-to-date descriptive model for decisions under risk. In accord with what we have said above, we think that it matters for normative adequacy if a factual deviation is systematic and intuitively plausible. In order to avoid dogmatic normativism, decision theory must say something on why people do not do what they ought to do, or it must be adjusted to the observed phenomena. With regard to the St. Petersburg paradox, both strategies fail. It seems mistaken simply to dismiss common intuition as irrational, and yet decision theory cannot be modified so as to produce the desired result. This suggests that we must look for a different standard of rationality operating in the St. Petersburg game.

4 Virtue and Emotion in Risk Assessment

The St. Petersburg paradox shows that our tendency to overweight extreme, but unlikely outcomes, as observed by Kahneman and Tversky, turns into the opposite at some point. We tend to ignore even the most extreme possible outcome if it is only unlikely enough. Outcomes with a probability below a certain threshold value are ignored, whether or not decision theory tells us that the magnitude of the outcome overweighs its low probability. In other words, decision theory, be it in the standard or in a more sophisticated version, is ignored when we play the St. Petersburg game.

Why should we assume that this indifference to decision-theoretic wisdom is rational? Why shouldn't we instead buy, e. g., the psychological explanation that our brains get overloaded so that the risk of extreme events is discounted because the probability is too low to evaluate intuitively? We offer three reasons: First, in the St. Petersburg game the systematic deviation observed from the decision-theoretic standard of rationality is of a remarkably high frequency. No reasonable person does, or would choose to pay any finite entrance fee to enter the game so that the deviation is 100%. Secondly, the confidence in this "intuitive" choice is unshakable. This distinguishes the St. Petersburg paradox from many other anomalies in which people adjust their choices to the decision-theoretic standard of rationality when they learn what this standard prescribes. A prime example is Tversky's and Kahneman's (1981) "Asian disease problem", in which people even feel ashamed of having violated this standard. By clear contrast, people do not revise their intuitive choice in the St. Petersburg paradox: the confidence in intuition remains unshattered

by the discrepancy between intuitive and apparently rational choice. Thirdly, though not in terms of decision theory, the intuitive choice can be justified. It would clearly be foolish to stake a lot of money on an outcome that is possible but extremely unlikely, even if the outcome is very, very large. It would be foolish because it is extremely likely that the event will never happen and thus never play any role in our life. The St. Petersburg game can make a millionaire – but the probability that it will not is one million times higher. As Weirich (1984, 198) puts it, “there is some number of birds in hand worth more than any number of birds in the bush”.

Yet, if not some standard proposed by decision theory, which other standard of rationality may be at work here? Speaking of a “standard” of rationality is misleading in so far as it suggests that decision making in the St. Petersburg game is done by a conscious process of inference. It suggests that the rational agent uses some algorithm, similar to the algorithm of decision theory, in order to make up his mind, while it seems that the decision is reached in a different way. In personal experience the decision presents itself as immediate, and that is: not as the product of an inference drawn by the agent. Furthermore, it is not only non-inferential in the phenomenological sense but also in the epistemic sense. If we are challenged about our choice in the St. Petersburg game, there is no valid inferential justification available to which we might “retreat”. The only valid inferential justification available is provided by decision theory, according to which one ought to play.

In the case of the St. Petersburg paradox, we claim, decision making is non-inferential because it is the product of virtue. What we have loosely described as an “intuitive” choice so far actually is a virtuous choice. Understood as virtue, risk assessment is a person’s capability to “see” risk in the right way, where seeing includes being motivated to act accordingly: it is *φρόνησις* (phronesis or *practical* wisdom), not just *σοφία* (sophia or *intellectual* wisdom; on this kind of seeing see in more detail Döring 2007, 2009; see also McDowell 1998; McNaughton 1988; Goldie 2007). Were we to follow decision-theoretic advice in the St. Petersburg game, this would deprive us of all means to pursue other goals. The utility maximising strategy trades approximately certain bankruptcy for approximately no chance of making a fortune. Rather than adopting this foolish strategy, it is clearly better to focus one’s thought and action on options that are likely to have an impact on one’s life. Recognising this is a matter of immediate insight. What distinguishes the virtuous agent, who does not care about decision theory in the St. Petersburg game, is his capacity to get his priorities right. More precisely, it is his capacity to attach appropriate importance to possible events in light of their probability, or to assess risk appropriately. This capacity has nothing to do with drawing inferences in the first place, although, as an expression of practical wisdom, it will influence the way in which its possessor draws inferences, and also which inferences he draws.

One might wonder how the capability in question could possibly be a virtue if there is no obvious name we can give for it, as we can do for classical virtues such as honesty or generosity. On our view, it suffices that virtuous risk assessment can be described as the mean between two “vices”, just as virtue is defined by Aristotle. The virtuous risk assessor would neither be a bold gambler nor a wretched coward, but choose an option in between these vicious extremes. Due to the excellence of his

character, he would balance caution and courage in an ideal way. This virtue account of risk assessment has a firm place in our everyday practice. We call someone a “coward”, or “overcautious”, if he rejects promising options with a risk that others would tolerate. Conversely, someone is classified as a “gambler”, or “daredevil”, if he runs options that others would find too risky. In both cases the reference point is virtuous risk assessment. In our everyday practice, we judge the appropriateness of a decision under risk relative to what the virtuous person, the *φρόνιμος*, would do.

Virtuous risk assessment, in its turn, involves having appropriate emotions. This fits the findings of Slovic (2004; see chapter “If I look at the Mass I Will Never Act: Psychic Numbing and Genocide”, this volume), who distinguishes between “risk as analysis” and “risk as feeling”. According to Slovic, there are two fundamental ways in which human beings assess risk. Risk as analysis uses algorithms and normative rules, such as probability calculus and formal logic. It is relatively slow, effortful, and requires conscious control. By contrast, risk as feeling is experiential, intuitive, fast, and mostly automatic. We think that, at the empirical level, virtuous risk assessment corresponds *cum grano salis* to Slovic’s risk as feeling, although we would prefer “emotion” to “feeling” in order to emphasise the cognitive function of the mental states in question. In relation to other mental states and actions the emotions play a dual role. On the one hand, they have motivational force. On the other hand, they can justify other states and actions, and they do so in the non-inferential way of perception. As we have argued elsewhere (Döring 2007, 2009; cf. also Tappolet 2000; Johnston 2001), at least paradigm cases of emotions may be characterised as perceptions of value, taking value in a broad sense. Because of their dual role, the emotions are natural candidates to be associated with virtue (see, e. g., Goldie 2007).

Again, it might be objected that it is hard to see which particular emotion should be involved in making our decision in the St. Petersburg game. Again, we reply that the emotional character of this decision can be seen from its place between two extremes. The coward rejects promising options with a risk because he is too fearful. Conversely, the gambler chooses too risky options because he is not fearful enough to assess the risk appropriately. Emotions are also involved at the metalevel of assessing risk assessment, as can again be seen most clearly in cases of deviations from virtuous risk assessment. Both cowardice and daring raise intense negative emotions, such as resentment, indignation, shame, or disgust.

5 The Inverted St. Petersburg Paradox as a Model of Risky Technologies

So far we have treated the St. Petersburg paradox as a logical challenge to the normative theory of rational choice. Yet, contrary to the predominant view, we also believe that there are decision situations in real life which can be adequately modelled by the St. Petersburg game. In other words, we claim that the systematic deviation from decision-theoretic rational choice and its rational explanation as virtuous choice are

of empirical salience. In particular, the St. Petersburg game can be applied to the assessment of risky technologies. To show this, it will prove useful to introduce the inversion of the game.

Rather than being rewarded with a prize reciprocal of probability, the gambler is punished with a fine of the same amount. Accordingly, the decision he has to make is not about an entrance fee, but about an “insurance premium”. In the inverted St. Petersburg game, the question is how much one is prepared to pay in order to avoid the game (Table 2). Here is the corresponding table:

Table 2 The inverted St. Petersburg game

n	P(n)	Fine (\$)	Expected payoff (\$)	Cumulative expected payoff
1	1/2	-2	-1	-1
2	1/4	-4	-1	-2
3	1/8	-8	-1	-3
4	1/16	-16	-1	-4
5	1/32	-32	-1	-5
6	1/64	-64	-1	-6
7	1/128	-128	-1	-7
8	1/256	-256	-1	-8
9	1/512	-512	-1	-9
10	1/1024	-1024	-1	-10

Again, the expected payoff of the whole game adds up to infinity, although in this case to minus (rather than plus) infinity. That is, in the inverted St. Petersburg game any finite premium is smaller (in absolute terms) than the expected value of the game so that, just as in the standard version of the game, decision theory commits one to play at all costs. Again, this contradicts what we immediately see as rational.

Maybe, our assessment of the game is not perfectly symmetric in both cases. Maybe, we would invest more to avoid the inverted St. Petersburg game than we would pay to enter the standard St. Petersburg game. This fits Kahneman’s and Tversky’s observation about “loss aversion”, which we already described. People have different risk attitudes towards gains (outcomes above the reference point) and losses (outcomes below the reference point) and care generally more about potential losses than potential gains. But this slight difference in subjective value between the standard and the inverted game does not affect the paradox. In both cases the verdict is that decision theory fails to resolve the St. Petersburg paradox. In both cases the rational decision is, in our view, the product of virtue.

One might wonder whether it makes a difference that in the inverted game one seems in a way “forced” to play, whereas in the standard game one is free whether or not to bother playing it. This question might arise because games are often presented as gambles, and we presume that the gambler enters the casino voluntarily. But in decision theory games are understood as models of the structure of all kinds of real life decision situations, and in reality one does not only make free choices from a status quo which is certain. Real life choices concern the avoidance of potential

losses no less than the bringing about of potential gains. Often, one has to make up one's mind about how much one would be willing to pay in order to insure against the risk of a large loss. Put more generally, one often has to decide how much certain utility one is willing to sacrifice to insure against the risk of a large disutility. We claim that decision situations of this kind are characteristic of risky technologies. More precisely, decision situations structured like the inverted St. Petersburg game are characteristic of risky technologies. Often we must decide whether or not to use a risky technology. If we decide not to, we sacrifice the utility which would have emerged from using it. The sacrificed utility corresponds to the insurance premium in the inverted St. Petersburg game. Conversely, to enjoy the benefits of a risky technology is to refuse to pay the insurance premium. The large loss or large disutility, modeled as the fine, is the outcome of a "beyond-design-basis accident" in many risky technologies. Beyond-design-basis accident is a concept employed in risk assessment of nuclear power plants. It refers to an accident which causes more harm than a "design-basis accident", i. e., a postulated accident that a technology must be designed and built to withstand.

We may here consider nuclear power as the classical example of a risky technology which involves a choice to be modeled on the inverted St. Petersburg game. We shall leave aside the multiplicity of more or less manageable malfunctions and costs, which require a different framework. Other things being equal, the decision whether or not to run nuclear power plants can be described as a choice between the safe option of keeping the status quo and the option of enjoying the benefits of nuclear power, such as cheaper power, jobs, pure research, or technical progress. The risk is a beyond-design-basis accident with an expected disutility which is more or less infinity – serious environmental contamination, injury, death – at a probability which is extremely low – estimates vary between 1/400,000 and 1/30,000 per year and plant for a meltdown and a tiny fraction of this for a subsequent radioactivity release (given German security standards).

Now, the choice whether or not to use nuclear power is the choice of a group. Thus one might object that our example is not in accord with the principle of methodological individualism: the St. Petersburg game must be offered to an individual, and not to a group. Rather than elaborating how decision theory and collective choice relate, we simply give another example. Consider driving a car as an instance of the use of a risky technology. In simplified terms, the decision whether or not to use the car is a choice between the safe option of staying at home and the risky option of enjoying the comforts and convenience of driving. Again, the risk is a beyond-design-basis accident with an expected disutility which is more or less infinity – death – at a probability which is extremely low – statistically about 1/14,000 per year in Germany; 1/7,700 in the US.

To say that the assessment of risk in these examples is analogous to virtuous risk assessment in the (inverted) St. Petersburg game is not to say that the risks of nuclear power or driving or whatever technology should be ignored. Our point is not that, since it is rational to ignore extreme, but highly unlikely outcomes in the St. Petersburg game, we should equally ignore extreme, but highly unlikely outcomes in assessing risky technologies. All we are claiming is that decision theory

fails in these cases, and must be substituted or supplemented by virtue. Virtue is appropriate for risk assessment in real life decision situations showing the structure of the (inverted) St. Petersburg game, without this implying any definite decision. The threshold value below which we ignore risk may greatly vary with the magnitude of the outcome.

In any case, risk assessment as virtue is not merely concerned with the two variables probability and outcome measured in utility. These are the parameters of consequentialist decision theory. Decision theory and virtue represent two fundamentally different approaches to value. Decision theory is a sophisticated method to operationalise the idea that it is the consequences of actions that matter. The value of an action is assessed solely in terms of its consequences and consists in the contribution of those consequences to people's utility, where utility is based upon the satisfaction of subjective preferences. Decision theory is a tool that helps us to evaluate possible outcomes with the highest precision possible. The St. Petersburg paradox shows that this tool needs to be supplemented by virtue even in idealised cases in which only the consequentialist parameters of rational choice are considered. When we turn to real life decision situations, many more aspects play a role which cannot be reduced to probability and degrees of preference satisfaction, and which are yet accounted for by virtue.

By contrast with consequentialist decision theory, virtue theory is concerned with value in terms of excellence of character. The leading question is what kind of person one should be. Of course, the answer to this question depends on what we care about, on what we think is valuable, so as to enhance the quality of human life. This is where the emotions come in. In our emotional reactions we evaluate things in light of what we care about (see also Helm 2001). To fear a risk, or to be ashamed and regret that one has (not) taken it, is not simply to have a subjective preference for or against a certain thing. It is to evaluate that thing in a certain way, and this evaluation takes the form of the perception of a value. As perceptions of value, the emotions differ from subjective preferences in that they are subject to a standard of appropriateness. The virtuous person experiences appropriate emotions and can thus immediately see what is valuable (Döring 2003, 2007, 2009; see also Döring and Peacocke 2002; Tappolet 2000).

By claiming that rational risk assessment in the St. Petersburg game depends on virtue and emotion, we do not mean to say that the decision-theoretic standard of rationality is to be abandoned. We think that both methods of practical reasoning are valid for evaluating risk. As a game with infinite expected value the St. Petersburg game clearly is a limiting case. For games with moderate outcomes and moderate risk decision theory is of great benefit in the assessment of risk. Yet, as we will now elaborate to conclude our argument, even here decision theory cannot do without virtue.

As we have seen, risk assessment as virtue has a firm place in our everyday practice. Therefore, it does not come as a surprise that the decision-theoretic description of our attitudes towards risk is not in accordance with everyday language. Again, this can be demonstrated at the St. Petersburg game. In decision theory, risk aversion is defined as the preference of a certain option over a risky option with a

higher expected value. Thus paying less than \$25 for the standard St. Petersburg game, which has an infinite expected value, is described as risk-averse behaviour. By contrast, paying less than \$25 to avoid the inverted St. Petersburg game is seen as risk-seeking behaviour, since one refuses to spend even \$25 to insure against an infinitely high risk. This attribution of attitudes does not fit the use of the terms “risk-averse” and “risk-seeking” in ordinary language. In ordinary language, someone is called “risk-averse” if he insures against risks that others would tolerate. Someone is classified as “risk-seeking” if he enters a game that others would find too risky. That is, to be risk-averse or risk-seeking does not mean to prefer less or more risk than expected value. The reference point is not risk-neutrality but virtuous risk assessment. This normative account of attitude towards risk is not taken into account by decision theory. If decision theory is to contribute to risk assessment in a normatively adequate way it should at least restrict risk preferences so as to render the “vicious” extremes inadmissible. Preferences implying daring or cowardice should be excluded from the rational domain, as it is exemplified in Fig. 4 below.¹

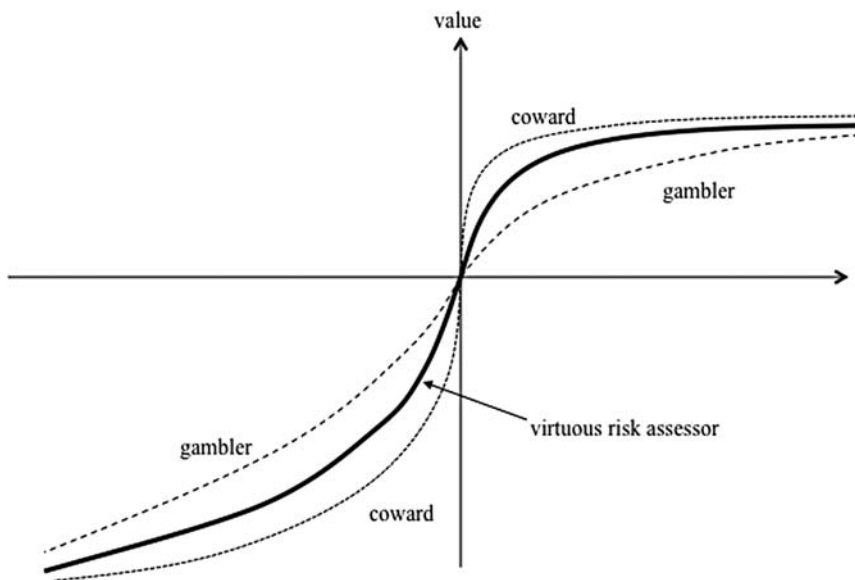


Fig. 4 Value function showing the “virtuous” reference attitude towards risk and the “vicious” limits of the admissible corridor

¹It has not been the point of this paper to specify the conditions under which non-inferential, emotionally grounded risk assessment is appropriate and thus in fact virtuous. That issue is a special case of the general issue of how value and evaluative properties (such as being risky) could be known, if there is such a thing as knowledge of values in the first place. The claim that emotions are subject to a standard of appropriateness is obviously compatible with (some sort of) value realism, but it does not entail it; one could still insist that there are no evaluative properties out there. However, many resist scepticism about the existence of values today, and so do we. Currently available options of value realism include, e. g., Johnston’s (2001) “detectivism”, Helm’s (2001) “holism of import”, and McDowell’s (1998) “sensibility theory” as a special case of a “buck-passing-account” (Scanlon 1998, pp. 95–97). It will be a matter of future work to make a rational decision between the available options – or to choose yet another one.

References

- Arrow, K. 1964. *Social Choice and Individual Values*. New Haven: Yale University Press.
- Bernoulli, D. 1738/1954. Exposition of a new theory on the measurement of risk. *Econometrica* 22: 23–36.
- Blavatsky, P. R. 2005. Back to the St. Petersburg paradox? *Management Science* 51(4): 677–678.
- Döring, S., and C., Peacocke. 2002. Handlungen, Gründe und Emotionen. In *Die Moralität der Gefühle*. S. Döring, and V. Mayer, eds., Berlin: Akademie Verlag.
- Döring, S. 2003. Explaining action by emotion. *The Philosophical Quarterly* 53(211): 214–230.
- Döring, S. 2007. Seeing what to do: Affective perception and rational motivation. *Dialectica* 61(3): 363–394.
- Döring, S. 2009. Why be emotional? In *Oxford Handbook of the Philosophy of Emotion*. Oxford: Oxford University Press (in print).
- Friedman, M., and L. J., Savage. 1948. Utility analysis of choices involving risk. *Journal of Political Economy* 56(4): 279–304.
- Goldie, P. 2007. Seeing what is the kind thing to do: Perception and emotion in morality. *Dialectica* 61(3): 347–361.
- Gustason, W. 1994. *Reasoning from Evidence*. New York: Macmillan College Publishing Company.
- Hacking, I. 1980. Strange expectations. *Philosophy of Science* 47(4): 562–567.
- Hardin, R. 1982. *Collective Action*. Baltimore: The John Hopkins University Press.
- Helm, B. 2001. *Emotional Reason*. Cambridge: Cambridge University Press.
- Jeffrey, R. C. 1983. *The Logic of Decision*. 2nd ed., Chicago: University of Chicago Press.
- Johnston, M. 2001. The authority of affect. *Philosophy and Phenomenological Research* 63(1): 181–214.
- Kahneman, D., and A., Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2): 263–292.
- Markowitz, H. M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New Jersey: John Wiley & Sons.
- Martin, R. 1998/2004. The St. Petersburg paradox. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/paradox-stpetersburg/>. Accessed May 2006.
- McDowell, J. 1998. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- McNaughton, D. 1988. *Moral Vision. An Introduction to Ethics*. Oxford: Blackwell.
- Menger, K. 1934. The role of uncertainty in economics. In *Essays in Mathematical Economics in Honor of Oscar Morgenstern*. M. Shubik, ed., Princeton: Princeton University Press.
- Pratt, J. W. 1964. Risk aversion in the small and in the large. *Econometrica* 32: 122–136.
- Rabin, M. 2000. Diminishing marginal utility of wealth cannot explain risk aversion. *Institute of Business and Economic Research Paper*, E00-287. <http://repositories.cdlib.org/iber/econ/E00-287>. Accessed May 2006.
- Resnick, M. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis: University of Minnesota Press.
- Rieger, M. O., and M., Wang. 2004. Cumulative prospect theory and the St. Petersburg paradox. *Sonderforschungsbereich 504 Working Paper* 04(28): Mannheim.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Harvard: Belknap.
- Slovic, P. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24(2): 1–12.
- Tappolet, C. 2000. *Émotions et valeurs*. Paris: Presses Universitaires De France.
- Tversky, A., and D., Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–458.
- Tversky, A., and D., Kahnemann. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives* 5(1): 193–206.
- Tversky, A., and D., Kahnemann. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4): 297–323.
- Weirich, P. 1984. The St. Petersburg gamble and risk. *Theory and Decision* 17(2): 193–202.

Emotions and Judgments About Risk

Robert C. Roberts

1 Introduction

This paper has six parts. First I outline a view of the relation between emotions and judgments that is suggested by an account of the nature of emotions that I have defended elsewhere: the idea that emotions are concern-based construals (Roberts 2003, Chapter 2). The relation between emotions and judgments is very much like that between ordinary perceptions and judgments. The way things appear to us perceptually is a standard basis for judging them to be as they appear; but the perceptual appearance of things can also lead us to make false judgments. So also with the perceptions that we call emotions. In the second part I discuss some epistemic advantages and liabilities that emotions lend to judgments. In the third, I clarify the concept of a risk, and thus of a judgment of risk. Fourth, I propose that the emotion types most pertinent to risk-judgments are members of the fear family. Fifth, I apply the points I have made in the earlier parts to some empirical findings on emotions and judgments concerning risk. Some of these findings bear on specifically moral judgments concerning risky technologies. I conclude the paper with a few brief comments about moral judgment.

Emotions make unique contributions to the epistemic excellence of some kinds of judgments, especially value judgments and practical judgments. But just because of their epistemic potency, emotions can also degrade our judgments. Given this dual truth, a normative account of emotions and judgments must bring in the concept of virtues. Some virtues are dispositions to experience correct emotions, but dispositions to experience correct emotions cannot carry all of the normative load; it is psychologically unrealistic to think that a person can be formed in such a way as to have only judgment-enhancing emotions, and never judgment-degrading ones. Some of our virtues will be dispositions to respond with correct emotions, and others will need to be ones that enable us to manage or transcend our emotions in the

R.C. Roberts (✉)

Department of Philosophy, Baylor University, Waco, TX, USA

e-mail: Robert_Roberts@Baylor.edu

interest of correct judgments. So we will need at least two kinds of virtues to maximize truth and minimize error in our judgments that are conditioned or influenced by emotions.

I assume that judgments are truth claims, and that we are all makers of judgments, many of which are affected, for good or for ill, by our emotions. This assumption implies that we are not engaged in a merely descriptive enterprise concerning how emotions are related to judgments, but are interested in putting ourselves in the best possible position to make the best possible judgments. Thus if individuals differ from one another in the excellence of their judgments, and we are in a position to improve ourselves with respect to our emotions and our judgments, then in studying this subject, we are engaged in a personal and normative enterprise. I hope that my remarks will bear on this interest and engagement.

2 Emotions and Judgments

Emotions are related to judgments in much the way sense perceptions are related to judgments. I look in the cupboard and my eye falls on a sack of flour. I form, quite automatically, the judgment that *a sack of flour is in the cupboard*. This is the typical case. Sense perception gives rise to belief. But perceptions are not themselves beliefs, and the connection between the two is not necessary; the beliefs that perceptions naturally engender can be stopped. Fairly early in life, we learn to be critical of our perceptions. The Mueller-Lyer Illusion is a striking example. We draw two parallel lines, measuring them to make sure they are exactly the same length. Then we add the little tails, and the line with the outward-pointing tails looks longer than the one with the inward-pointing tails. We *perceive* the one line as longer than the other, but for good reasons we confidently and simultaneously refuse the *judgment* that the one line is longer than the other. The one line continues to *appear* to us longer than the other; but we withhold *assent* from the appearance, and thus refrain from judging the lines to be of different lengths.

The purely *sensory* input contributed by the parallel lines in the Mueller-Lyer illusion is not enough to promote the judgment that the lines differ in length; were we to measure the lines projected onto the retina, they would presumably differ no more in length than do the lines on the paper; yet the lines *appear* to differ in length. This combination of facts suggests that the appearance of the lines as differing in length is not sensory in the strictest sense, but is a *construction upon* the sensory. It is what I call a “construal.” This construal is illusory; if one were to form, on the basis of it, the judgment that the lines are of unequal length, the judgment would be false. Other drawings elicit construals that are not illusions, for example, the duck-rabbit and the old woman/young woman figure. Each of these yields two quite different possible perceptions, from one set of sensory data. The data underdetermine the perceptions, showing that the perception is a construction of the subject. A construal, in my special sense, is a perception, but it is never purely sensory; that is the point in calling it a construal (construction). It is what the Greek

Stoics called a *phantasia*, and in Aristotle it falls under the concept of *aisthêsis* (perception). So we have construals that are false, and construals that are neither false nor true. Still other construals are positively veridical. For example, when a hospital pathologist looks at a slide of tissue in the microscope and sees melanoma, thus forming the judgment that the tissue is affected by melanoma, the perception is a construction of the purely sensory data, a construction based in the scientific concepts and practices of tissue pathology; nevertheless, both the construal and the corresponding judgment may well be true.

Note also that the perception of the Mueller-Lyer drawing has a propositional structure; it makes a sort of “claim.” It “says” something to the effect of, *the upper line is longer than the lower line*. If the perception lacked this structure, then it couldn’t be contradicted by the judgment that the two lines are the same length. It is typical of construals, as I understand the term, that they have a propositional structure. The construals of the old woman/young woman figure would “say” *this is a picture of an old woman with such-and-such features* or *this is a picture of a young woman*, etc.; the pathologist’s construal says *this is melanoma*.

The illusory construal of the Mueller-Lyer drawing seems to be strongly predisposed by human nature; people do not have to learn to see it this way, and it is very difficult to learn not to see it this way. By contrast, one has to *learn* to see some of the gestalt drawings as they predispose. For example, I couldn’t see the young woman in the figure until my wife trained me to do so, despite the fact that the drawing is carefully and successfully contrived to predispose quite particular perceptions – the young woman and the old. And one goes to school to learn to see melanoma on a microscope slide.

In many contexts of life, an important part of learning to make accurate judgments about things is learning to perceive in a way that yields correct judgments. For another example, an auto mechanic learns to hear things going on in a running engine that untrained people don’t hear, and his perceptions make it possible for him to come to reliably correct judgments about the state of the engine.

The above points about sensory construals and their relation to judgments can be made also about emotions. But emotions are a special kind of construal. I call them “concern-based construals.” It is very common to form judgments on the basis of or in coordination with our emotions. If someone fears walking down a certain unlit street at night, it will be quite natural for him to judge that the street is dangerous. The fear may come upon him as he enters the street, without his having any other “information” about possible dangers than the “look” and the “feel” of the street. He finds himself afraid, in much the way that we find the lines of the Mueller-Lyer drawing looking different in length; and he spontaneously forms the judgment that the street is dangerous. But the connection between the emotion and the judgment here is no more necessary than in the case of the drawing. Someone might be convincingly informed that the street is safe, and still feel afraid as he walks along. In this case, he fears the street but does not judge it to be dangerous. It continues to have the “look” or “feel” of being dangerous, but the subject actively discounts this impression, just as one discounts the impression that the two lines are of different lengths.

I say that fear is a construal that, like other construals, goes beyond any sensory information that the feared situation may yield. It is a perceptual construction of the situation as containing a certain moderately high probability that something bad is going to happen. One difference between emotions and the construals that we have been considering so far is that many emotions are even more underdetermined by sensory data. One can become angry, or fearful, or hopeful, or joyous by merely thinking about situations in a certain way. One may or may not have quasi-sensory experience of the situation through mental images, although imagery enhances the probability of feeling the emotion. Still, the emotion is perception-like in that the situation “comes together” for one, with an immediacy analogous to sense-perception, in the terms characteristic of the emotion in question (*bad possibility* for fear; *good possibility* for hope; *culpable offense* for anger, etc.).

Another way such construals differ from non-emotional ones is that the situation is construed as impinging on one or another concern of the subject (in the case of fear and hope, the concern for wellbeing; in the case of anger, the concern for whatever is construed as offended against; etc.). In each case, the situation would not have the “look” that it has for the subject were he not concerned about it in the way he is concerned. This aspect of the construal gives it its “feel,” its sense of urgency, the vivid appearance of relevance to what is important, its evaluative character and its power to motivate.

Sense perceptions can yield *judgments* because they already have a propositional structure; the perception yields a judgment by virtue of the subject’s assenting attitude to what the perception “says.” I propose that in the same way, the concern-based impressions that we call emotions have a propositional structure: they present the situation to the subject as being a certain way, which can be expressed (more or less approximately) in a declarative sentence. They “assert” something about the situation. In my recent example, the emotion presents the situation of the walker in the street as containing an uncomfortable probability of something bad happening to him. It is as though the emotion says to him, in the language of impressions, *Something bad may happen to you here*. When the person makes a judgment corresponding to the emotion, it is as though he is agreeing with the impression; but he may dissent from it, despite the fact that the emotion is screaming in his ears, as it were, “Something bad may happen to you here,” urging him to flee the situation. A difference between emotions and mere sensory construals like the Mueller-Lyer illusion is that dissenting from the latter is easier because it is not complicated by the motivational element. As concern-based, emotional construals grip us more deeply, more insistently; they speak to our life, whereas the merely cognitive impressions speak only to our “intellect” (who cares, after all, whether the lines are the same length?!). So it typically takes more moral or quasi-moral maturity to disbelieve our emotions than it takes to disbelieve our eyes.

People differ in the accuracy and subtlety of their impressions. When we want to find out whether our engine is going to let us down in the next 5,000 km, we don’t trust the hearing of just anybody. We want an expert mechanic to listen to our engine. Training in all kinds of areas involves training in perception. Think of ear training in music, of bird watching, various kinds of microscopy, art history, expertise in

reading various kinds of meters and other scientific instruments, of stage direction, paleontology, etc. etc. Again, we see that perception has to be more than sensory; in each case, the non-expert's purely sensory impressions of the object may be the same as those of the expert, but he doesn't see or hear what the expert sees or hears. He doesn't know how to "read" (construe) the sensory data.

The idea that people can be more or less mature emotionally is a familiar one. Little children are sometimes afraid of the dark, and we are happy when they outgrow this disposition. Adults fear things that children don't fear — a fall in the stock market, cancer, cholesterol, alcohol. Sometimes adults fear such things more than they are worth fearing, and sometimes they don't fear them enough, or they fear them for the wrong reasons. Adults sometimes have phobias – tenacious dispositions to fear things that are not fearsome. Irrascible people get angry at things that don't warrant anger. People sometimes go to therapists for treatment of their phobia or irascibility. But we do not want people to be so formed as not to fear anything at all, nor ever to get angry, and often the therapist can help, either by advice that leads to better habits, or by some other way of affecting the disposition for the better. So there must be such a thing as correct emotional dispositions, and the aim of much moral and psychological education is to form our emotion dispositions, so that we see the world in *correct*, or at least *better*, evaluative terms.

Aristotle (1980, Book 4, Chapter 5, pp. 96–98) thinks that many of the virtues are dispositions to proper emotion. Thus the virtue he calls *praotês* (mildness, gentleness) is a disposition to get angry with just the right people, at just the right time, for just the right reasons, and not to get angry otherwise. I think we know people like this, or at least people who approach more closely getting it right about the objects of their anger; and we also know people whose anger is too much, at the wrong time, towards the wrong people. Another example of Aristotle's (1980, Book 3, Chapters 6, 7, and 9, pp. 63–72) is *andreia* (courage). It is a disposition to fear just what is genuinely fearsome, in just the degree to which it is fearsome, for just the reasons that warrant the fear, and so forth. His idea is that people's emotion dispositions can be trained to get their objects right. If emotions are perceptions of situations in their evaluative dimensions, and perceptions are proto-judgments or dispositions to form judgments, then it would appear that a person with the virtues will be something like an "expert" in making evaluative judgments of the kind in question (in the case of *praotês*, judgments involving offense; in the case of *andreia*, judgments regarding risks to himself).

3 Some Epistemic Advantages and Liabilities of Emotions

A number of epistemic criteria can be used to evaluate judgments. Here I will mean by a *judgment* a particular episode of assenting to, or being in an assenting attitude toward, a proposition. A judgment, in this sense, is always somebody's, at a given moment; it is not the sort of thing that can appear in a book, though somebody writing a book or reading one may make a judgment that corresponds to a formula that appears in the book. I will be considering judgments that are made, or can be

made, in conjunction with emotions, and will be asking whether, how, and to what extent, the association of the emotion with the judgment affects its epistemic value. I will now consider four criteria for evaluating judgments: Correctness, justification, experiential immediacy, and understanding. They are epistemic goods, or properties that a judgment can have, or a person can have in making a judgment.

3.1 *Correctness*

Other things being equal, correct judgments are better than false ones. Official Stoic doctrine is that judgments formed on the basis of emotions cannot be true. The Stoic's reason for this claim is that emotions are about situations that are indifferent with respect to value (neither really good nor really bad), but they claim goodness or badness. I have said that fear is the impression that it's pretty probable that something bad will happen. One fears things like failure, disease, accident, and death; one hopes for things like success, health, safety, and longevity. But Stoic doctrine says that success is not good and failure is not bad; death is not bad and longevity is not good (though all these things may be "preferred" or "not preferred"). So judgments that involve assent to the propositional content of emotions cannot be true.

On this point, Stoicism is bizarre. If, to the contrary, we hold that disease and failure can be genuinely bad for people, and if we agree that emotions are construals, then we will think that judgments based on fear may sometimes be correct. And the same will be true of most other emotions: anger, hope, joy, contrition, gratitude, compassion, disappointment, and many others. But we will also think, no doubt, that very often emotions construe their situational objects in ways that we should not assent to, because to do so would be to judge falsely. Stage fright is usually a false construal of the situation, as is hope in hopeless situations. So part of leading an epistemically responsible life is to try to have veridical emotions as much as possible, and to be able to withhold assent from the ones that are likely to be false.

3.2 *Justification*

Epistemic justification is a relationship that can obtain between a judging subject and his judgment, in virtue of some factor that justifies the subject in making the judgment (that is, assenting to the proposition). For example, a person might be justified in his judgment that his sailing vessel is in danger by his awareness that he is two miles from shore and the radio is predicting imminent rough weather, plus some other truths about a human being's being caught in rough weather in a small boat. The "factor" in this case is his *having the evidence* of the radio report and his distance from shore, and the other things.

A judgment can be true without being epistemically justified, and justified without being true. If the sailor judges himself to be in danger because the rabbit's foot he carries with him fell irretrievably in the water, then his judgment will not be

justified; and this will be so even if his judgment happens to be true because, unbeknownst to him, a dangerous storm is actually approaching. On the other hand, if the “radio report” is really a tape recording with which his buddies are playing a practical joke on him, he could be epistemically justified in his judgment without its being true.

The question before us is whether an emotion can ever be a “factor” in virtue of which a person is epistemically justified in making a certain judgment. Put in the terms of the account of emotions I am promoting, the question is whether a concern-based construal of a given situation can ever contribute to the epistemic justification of the judgment that the situation is as the emotion construes it, and if it can do so, under what conditions it can. I claim that it can, on the condition that the person having the emotion is emotionally fit – that his emotions of the type in question are generally indicators that the corresponding judgments are true. Continuing our sailing theme, imagine the following scenario. You, an inexperienced sailor, are on a sailing vessel and the water starts to get choppy. The boat is pitching this way and that, and water is coming over the deck. You begin to be afraid, but you are savvy enough about the nature of emotions and the state of your nautical ignorance that you don’t immediately judge yourself to be in danger. You *don’t* take your emotion as justifying the judgment that you are in danger. Instead, you watch the captain. As long as he seems to be unworried, you judge that things can’t be too bad. But if he begins to show signs of real anxiety, you (correctly) take this as evidence that the boat is in danger. You take *his* emotion as justifying the judgment.

You may think, Well, you may take his emotion as justification for your judgment, but this version of the true story doesn’t show that the *emotion* is what justifies the judgment. *He* is going not on his emotion, after all, but on evidence of other kinds: the weather, the size of the vessel, the distance from shore, etc. His emotion is just a *consequence* of sizing up the situation, not the essential way in which he sizes it up.

This is the question. Is the emotion incidental to his knowledge – a mere by-product of his judgments about the danger? Or is it related more essentially to these things? Where does the concept of *danger* come from, anyway? Is not *danger* a concept that locks certain features of the situation to certain human concerns for life and wellbeing? If the concerns for life and wellbeing were not in the picture, in what sense would the combination of situational factors constitute a danger? If this is so, perhaps emotions like fear and anxiety are the most basic way of apprehending situations as dangerous. Thus the person who was able to synthesize the human concern for wellbeing with the relevant factors of a given type – say, storms at sea – would be the best judge of dangers. And, on the account I am offering, he would be the best *judge* because he was the best *perceiver* of dangers at sea. He surveys the situation in light of the human concern for life and safety, and this is what he sees: a significant danger (or not, as the case may be). And because he (justifiably, wisely) trusts his danger-construals in this kind of context, he judges the situation to be significantly dangerous.

Emotions can also undermine justification. I have pictured you as pretty wise and self-controlled about the judgment that you are in danger from the storm at sea:

You realize that the emotions of the unfit are not reliable indicators of states of the world in their relations to human concerns, and you have enough self-possession to stop the judgment that would arise normally from your concern-based construal as the boat you are on rises and plunges in the mounting waves. But what if you are less emotionally and epistemically mature? What if you tend to leap to every judgment that your emotions propose? In that case, your fear will not only fail to justify your judgment; it will positively *undermine* your justification. Your precipitous and uncontrolled acquiescence in forming judgments from emotions deprives them of, or reduces, their justification. When you make a judgment that arises out of *your* emotion, at least in cases relevantly similar to the one at hand, we have no positive reason to think that it is true. Unlike the captain, *you* are a better judge of such situations if you can prescind from your emotions and depend more on calculation, because your emotions tend to distort your judgment. We will see more about this in the next section. This is perhaps clearer in the case where your judgment is false. The skipper is allowed a certain automaticity in the formation of his judgments of dangers at sea on the basis of his emotions; but you, being inexperienced, are not allowed this easy transition. Thus we see that justification by emotions is a highly person-centered and virtue-indexed matter. Whether emotions justify or undermine the justification of the corresponding judgments depends heavily on the capacities and concerns of the person who is having the emotion.

If a judgment is evaluative – as judgments about dangers and offenses necessarily are – then valuation has to be somehow elemental to the judgment, and emotion would seem to be an important way, if not the primary way, of getting the value dimension into the judgment. On the view I am proposing, the value dimension of the judgment would enter via the basis in concern – things have value for us, negative or positive, by way of our being concerned about them, of our caring for them, of our wanting them or wanting to avoid them, of our being attached to them and thus wanting things with respect to them. This is not to say that values are subjective, but it is to say that in this primary way they are subjectively accessed or known. Perhaps some values are dictated by human nature: there are such things as proper and improper functioning for human beings, given the way we are constituted. There are ways of being healthy or sick that are determined by the kind of beings we are. Some of these ways are social. For example, for a human society, justice is a way of being healthy. If so, then one's value judgments, to be true, need to track these values; but the way we track them is to be concerned about them – disposed to seek the good and avoid the bad, to rejoice in the good and hope for it, and to find the bad repugnant and fear-inspiring. And this point leads me to the next criterion of epistemic excellence.

3.3 Experiential Immediacy

Some knowledge is more “intimate” than other knowledge, more a matter of direct acquaintance with the object. Consider the person with prosopagnosia. Most of us

recognize the faces of people we know by the “look” of the face, but people with prosopagnosia lack this power of immediate acquaintance with “face-looks.” They see faces as agglomerations of face-parts rather than as wholes with a characteristic look. People with this deficit do sometimes know whom they’re looking at, and they do so by inference: George is the one with the large bent nose, and Susan has a mole under her right eye. But despite these folks’ justified true judgments regarding the identity of their associates, they are missing something epistemically, and what they are missing is not just full reliability in their power to identify people by their faces. They are missing the experience of how George and Susan *look*. The color-blind person may have perfectly justified knowledge that the traffic light before him is showing green (he may know this by knowing that the green light is in the bottom position); still, he would be in an even better epistemic position with respect to this judgment were he able to see the green – as green – for himself. Autistic people are typically deficient at seeing things like the glory of a sunset; they can know that a sunset is glorious (by way of testimony), but they can’t see the glory for themselves. Much of the knowledge that we seek spontaneously and enthusiastically has to some extent this character of experiential acquaintance.

Emotions give access to this kind of knowledge of value judgments. If anger is perception-like in the way I have proposed, the person who gets angry about an injustice has an intimate acquaintance with the badness of the injustice, an acquaintance that is lacking to the person who merely judges, with justification, that the injustice has been perpetrated. The person who becomes anxious about a dangerous choice has an immediate grasp of the riskiness in the situation (*evil-in-potentiality*) that is lacking to one who merely makes a justified true judgment that the situation is risky. The person who experiences joy over the healthy birth of a child “sees” the goodness in the event better than one who merely judges that the event is good. I do not want to say, with the “internalists,” that the people who make only the “cooler” judgments in these cases lack evaluative knowledge, or lack justification in their judgments. The concept of knowledge is broad enough to encompass these cool cases. But the people who make only the cooler judgments do lack a certain important *kind* of knowledge, and the kind they miss is a deeply evaluative (in moral cases, moral) knowledge. It should be clear how, on my view, emotions supply this kind of knowledge. They are concern-based *perceptions* of situations in their evaluative dimensions.

I will have a bit to say, under the rubric of understanding, about how the fact that emotions give us experiential acquaintance with a situation as having a certain character can also create epistemic liabilities.

3.4 Understanding

The subject of a judgment is better off epistemically if he understands the judgment than if not; and the better he understands it, the better off he is epistemically. Understanding is primarily a matter of grasping, creating, and/or being able to

follow out, *connections* in a series or system or array (for example, understanding a sentence syntactically; understanding a narrative or musical composition; understanding a theory or explanation; understanding a map). One cannot form a judgment at all without understanding, to some extent at least, the proposition to which one gives assent; and this involves grasping connections between parts of the proposition (subject and predicate). The “seeing” of a figure in a gestalt drawing is a perceptual *ordering* of the lines and patches, and thus a species of understanding. It is a way of “connecting the lines,” so to speak. I have suggested that emotions are a kind of construal, in which the connections among elements in a situation are pulled together in an order characteristic of the emotion-type (fear, anger, hope, contempt, etc.), and the situation is connected to oneself via one’s concerns. This pulling together of the situational elements is a sense-making presentation of the situation. The perceptual immediacy characteristic of emotions (and not characteristic of all judgments) adds a dimension of grasping connections that is not there in the merely “intellectual” understanding of the situation. In an emotion, the *significance* of a situation *comes home*.

Damasio (1994) describes a patient, whom he calls Elliott, with profound damage to the pre-frontal cortex. Elliott tests normal on several IQ and personality inventories, but is almost completely unable to experience emotions (he can become momentarily angry, and his sense of humor seems to remain intact). His test-performance suggests that he can make judgments of the form *I am in danger of injury or death*, but being afraid is no longer in his mental repertoire. Presumably he could still make the judgment in circumstances in which it would be true. Further, he need not be without justification for his belief; perhaps he bases it on good statistical evidence concerning the dangers attending the kind of activity he engages in. But we might justifiably think that his failure to feel fear means that he doesn’t understand the judgment very well. So there’s something defective about the way he holds the judgment. We might say that he doesn’t appreciate the connection to his own life, and also doesn’t grasp very well the significance of the combination that is expressed in the proposition. (He can *make* the connections, in the sense of *explaining* them; but he lacks a certain kind of intelligence in his grasp of them. So in one sense he does understand the situation; but he does not understand it in the perceptual way that I am here trying to identify.)

Emotions’ character as potential acquaintance-knowledge can also blind us epistemically. Aristotle noted the common human experience of knowing that something is good for us – say, remaining sexually faithful to one’s spouse or staying off the booze or confronting our teenage children about their dangerous style of life – yet doing the opposite. The agent is not stupid or ill informed, yet he *acts* stupid and ill informed. Aristotle calls this phenomenon *akrasia* (incontinence, weakness of will). His solution to the puzzle is to distinguish two kinds of knowledge. There is dispositional knowledge, and the akratic person has plenty of this; it is in the dispositional sense that he knows what is best to do. But there is also a more direct *seeing in the particular situation* what is best to do and seeing it *as* the best; this is the sense in which he does *not* know what is best to do. And Aristotle says that *akrasia* occurs when a person is blinded by passion (Aristotle 1980, Book 7, Chapter 3).

If a passion or emotion is a more or less vivid perception of something as good or bad, attractive or repellent, then we can see how an emotion, by its very character as potential knowledge, is capable of blinding us. In the moment when the alcoholic contemplates the drink that is to be his downfall, it looks wonderfully good and attractive despite the fact that he knows in another sense that it is not, and his knowledge of the truth is overwhelmed by the compelling appearance of the drink's "goodness." When the father is about to have that serious talk with his teenage son, which he knows to be necessary for both his and the boy's wellbeing, his fear makes the action appear so threatening that over all, the situation looks more like a disaster about to happen than a golden opportunity. And in this blindness to the true character of the situation, the failure to see is at the same time a temporary confusion – a failure to understand, to see how the elements of the situation stand to one another.

4 Judgments About Risk

What is a risk? I propose that a risk is *a relatively high probability of harm or loss*. In performing some actions, a person *takes* a risk, that is, he subjects something that is valuable to himself, or something that should be valuable to himself, to a fairly likely harm or loss. A risk should be distinguished from a danger or a hazard. An unprotected cliff in a public park is a hazard. A person who lets his small children play near such a cliff runs a risk of losing them because of the hazard, but the dangerous cliff is not itself a risk. Persons with a certain gene *have* a higher than average risk of developing macular degeneration, but they don't thereby *run* a risk. They run a risk of macular degeneration only if they do something to enhance the risk, such as smoke cigarettes. People risk their lives in war, their money on the stock market, the forest when they leave a campfire undoused, their health when they smoke, and other people's health when they build a polluting factory in a neighborhood.

These examples suggest a couple of more points about the concept of risk. As in the case of the polluting factory, the possible harm need not be to the risk-taker. But the possible harm needs to be to something that the risk-taker at least *ought* to care about. When we speak about the risk of swatting a mosquito, we are not referring to the probability of harming the mosquito, but to the probability of harming ourselves or something we care about (say, our crystal goblet on which the mosquito has perched). The harm to the mosquito is not the kind that typically grounds a risk for a human being, because it is not a harm to the agent or anything whose harm the agent ought to be concerned to avoid. By swatting at a mosquito I do not risk the mosquito's life, because I neither care, nor should care, about preserving it. By contrast, if I build a polluting factory in some people's neighborhood, I risk their harm whether or not I care about it – because I *should* care about it.

People typically take risks for some envisioned benefit: the objectives of war, possible gains on the stock market, the convenience of not having to seek water to douse the campfire, prospective profits from the factory. The envisioned "benefit" can be the avoidance of some other harm, say, of being enslaved by an enemy

force. A risk – even a high probability of a significant harm – can be justified if the envisioned benefit or the probability of some other harm is great enough. Thus a high-risk surgery can be justified by the great benefit envisioned, should it succeed, or by the harms associated with not doing it. But we might doubt whether the prospect of looking 10 years younger if it succeeds justifies a high-risk surgery, or the thrill of skiing down a treed slope at 50 mph justifies the risk.

The above remarks suggest a number of possible ways for judgments to go wrong or right when an agent considers performing a risky action:

- Failing to notice/noticing the possible harm
- Misconceiving/correctly conceiving the possible harm
- Under- or overestimating, or correctly estimating the severity of the possible harm
- Under- or overestimating, or correctly estimating the probability of the possible harm
- Over- or underestimating, or correctly estimating the possible benefit of the action
- Over- or underestimating, or correctly estimating the probability of the possible benefit

5 The Emotion Type(s) Relevant to Judgments of Risk

We have a variety of perceptual powers, and they are to some extent specialized for kinds of judgment. Color judgments are in the domain of vision. Judgments about harmonic intervals are in the domain of hearing. Flavor and odor judgments are in the domain of tasting and smelling. (It is very awkward, or at least round-about, to make color judgments with your ears or tonal judgments with your eyes.) Of course many perceptual judgments use more than one sensory faculty: a judgment about what vegetable is on one's plate may exploit a combination of nose, tongue, and eye. And some judgments seem to be minimally or only very generally derivative from any particular sensory mode. Examples would be mathematical judgments and judgments of risk. In both cases, eyes and ears are no doubt often used, but these do not seem to be as essential to the kind of judgment in question as those about colors and tonal intervals.

Emotions, as ways of perceiving situations with respect to their values, come in types that are specialized for kinds of values or kinds of things that have value (positive or negative). Gratitude is about gifts, anger is about other people's offenses, guilt is about one's own offenses, pride is about honor, shame is about dishonor, hope is about good prospects, and so forth. Which emotion type is about risk? I have claimed elsewhere that the fear-family of emotion types are about bad prospects, and I have said that fear more particularly is about a certain probability of bad prospects. (Dread, by contrast, is about inevitable bad prospects. For a discussion of the nature and kinds of fear, see Roberts 2003, pp. 193–202.) Other members of the fear family are panic, anxiety, and cautiousness. Fear and anxiety are not specific to action, though we are often afraid or anxious when we perform risky actions. Cautiousness is specific to action, even to risk-taking, but it is not exemplified in all cases of risk-taking. By the same token that fear is the way we perceive bad prospects, hope is how we perceive good ones. If risky actions are taken for possible benefits, then

hope will also be involved in the perception of the situations in which we take risks. For brevity's sake I will here concentrate on fear, and say only a little about hope.

One of the explanatory advantages of conceiving emotions as construals is that it makes them more than events of affect. If we think of fear as a construal, it is not just dumb aversion, but aversion toward a *kind* of situation (namely, one containing a bad prospect of a certain degree of probability) for *reasons*. If I fear a coming hurricane, I *conceive* it as coming, and as a hurricane, against a background of *understanding* that a hurricane is a force that may destroy life and property. Only if my fear contains such thought content can it motivate me to take appropriate action: to drive away from the hurricane, to board up my windows and move lighter property to a protected location, etc. Since fear contains some conception of the harm that is feared, and some rough estimates of the severity and probability of the harm, it can match or mismatch the actualities of the situation it is about. To the extent that it matches the actual possible harm and its actual severity and probability, it is a promising perceptual basis for correct judgment; to the extent that it misrepresents these actualities, it is a poor one. We can make symmetrical comments about the conception of the benefit and estimation of its goodness and probability (hope). See the list of variables at the end of the previous section. The second epistemic good that I discussed in the third section of this paper – justification – does not require that the emotion be correct, but for most purposes it requires that the *agent* be *disposed* to experience emotions that are correct. The other two goods – experiential immediacy and understanding – gain much of their value from the value of correctness.

6 Emotions and Judgments About Risk

I will now consider some empirical findings concerning decision-making under risk, in light of the four criteria for excellence in value judgments and the ways in which emotions affect the quality of such judgments for good and for ill.

6.1 Risk/Benefit Confounding

Paul Slovic and his colleagues have identified a judgment-formation pattern they call “risk/benefit confounding” (see Alhakami and Slovic 1994; Finucane et al. 2000). In many cases, high-risk actions have a high potential benefit, while low risk actions yield low benefits. Contrary to this pattern in reality, risk/benefit confounding is a disposition to pair the perception of high risk with perception of low benefit and perception of high benefit with perception of low risk. The explanation appears to be that when a person's positive affect toward the benefit dominates his awareness of the situation, the perception of its goodness “colors” the risk in falsely positive hues; or when one's negative affect toward the risk dominates one's awareness, the perception of its badness colors the benefit in falsely negative hues. Plausibly, this epistemic infelicity is due to the acquaintance character of the perception: in the glow of the benefit, the risk *looks* low; in the ominous light of the risk, the benefit

looks unattractive. Judgments based on perceptions distorted by risk/benefit confounding will have a pretty strong tendency to be incorrect. Of course, high benefit is *sometimes* paired with low risk, and high risk with low benefit. In these cases, the perception produced by the confusion will be correct, but for somebody whose judgment is suffering from risk/benefit confounding, the correct perception will be accidental – just a matter of luck that this unreliable epistemic mechanism alighted on a case that it fits. Thus a judgment based on this perception, though correct, would fall short of justification.

Studies show that risk/benefit confounding is very common among human beings (Alhakami and Slovic 1994), but they do not show it to be inevitable or incorrigible. Social scientists tend to deal in trends and averages, but the virtue epistemologist is interested in especially well-developed *individuals* and the potential that we all have for moral and epistemic *education*. Social scientists tend to study how people *are* in the sense of *tend* to be, while virtue epistemologists and ethicists study how people *ought* to be within the boundaries of what they *can* be. We have no reason to think risk/benefit confounding is more resistant to remedial education than other fallacies to which human beings are prone.

Melissa Finucane and her colleagues have identified a judgment-forming short-cut that they call the “affect heuristic”: instead of “weighing the pros and cons or retrieving from memory many relevant examples” (Finucane et al. 2000, p. 3) for comparison, people often make a snap judgment on the basis of the affective valence of some currently salient stimulus. For example, if someone has just been told the many advantages of nuclear-generated electricity and then is asked how hazardous he judges nuclear generating plants to be, he might leap from the positive affect that he feels thinking about an inexhaustible and abundant supply of low-cost electricity, to the conclusion that they are pretty safe.

Finucane and her colleagues tested for the affect heuristic by limiting the time they allowed subjects for making their risk-benefit judgments, and they found that risk-benefit confounding occurred more under time pressure than when subjects had more time to consider the options. This experimental limitation simulates a feature of everyday life: sometimes judgments have to be made quickly. Most people who are careful in their judgments realize they are more likely to make a *good* judgment if they take more time. If one does not have time, then snap judgment on the basis of affective impressions is the best one can do. But the data seem to suggest that many people rely on the affect heuristic even when they are not under time pressure – perhaps out of laziness. Careful reflection takes effort, and many people are not habituated to make the effort as a default mode, nor alert to the importance of making effort in situations that call for it (they tend not to discriminate such situations from others that do not call for it). Since real life presents both kinds of situations – ones in which judgments of risk need to be made quickly and ones that allow time for deliberation – the best epistemic agent will be one who can make relatively good snap judgments in some area of judgment-making, but who can discriminate pretty clearly between situations in which he is competent and situations in which his snap judgments are less competent and then, when given the opportunity for careful reflection, avails himself of it in the interest of truth and justification. This disposition might fall under the virtues of intellectual caution and epistemic

self-knowledge, both of which presuppose a love of knowledge (see Roberts and Wood 2007).

Epistemically well-trained people learn to distinguish clearly between risks and benefits, and especially when they have been apprised of the natural and widespread fallacy of risk/benefit confounding, can be dispositionally on their guard against it. But if they have been sufficiently trained in making risk judgments (especially within some particular area of concern), they may not *need* to be on their guard against it, because they may have learned to make excellent snap judgments – even affectively laden ones. In the popular book *Blink: The Power of Thinking Without Thinking*, Gladwell (2005) gives many illustrations of *trained* snap judgments, which in many cases are emotionally laden and are sometimes more reliable than their elaborately and laboriously deliberated counterparts. Finucane and her colleagues seem to buy into the widespread and popular emotion/cognition dichotomy. Gladwell’s book, even in its title, challenges the supposition that rationality is necessarily discursive; reason is sometimes perceptual and often emotional. That is also a corollary of the view of emotions as concern-based construals: as construals, emotions are perceptions constructed, often, of a complex background of thought. If Gladwell and I are right, the affect heuristic need not be associated with such disreputable routines as risk/benefit confounding; some emotions might incorporate the most careful and rigorous distinction between risk and benefit.

6.2 *Anxiety Insensitive to Probabilities*

Weber and Hsee compared the maximum buying prices that subjects were willing to pay for “investment options that differed in the probabilities with which gains or losses of different magnitude would be realized” and compared these reports with “the degree of worry . . . they would experience between the time they invested in the option and the time they would find out which outcome actually occurred” (Weber and Hsee 1998, as reported in Loewenstein et al. 2001). They found that the subjects’ self-predicted behavior was sensitive to variations in probabilities and outcome levels, but that self-predicted worry was much less sensitive to probabilities. This result suggests that if people were to make their investment decisions solely on the basis of their emotional perceptions of risk, they would make less rational, and thus in all probability worse, investment decisions than they do, and also that people in fact do not rely entirely on their emotional responses to investment opportunities, but use calculation (or something similar) to moderate and correct for the effect of such responses. Given the discrepancy between the two epistemic modes, this procedure seems eminently rational, and exemplifies part of the picture of the intellectually virtuous person: he is guided, in making judgments, *both* by his emotions, *and* by emotion-transcending or -bypassing calculation.

But anxiety is already *somewhat* sensitive to probabilities, and emotions can be educated. We might expect that seasoned investors’ anxieties would correspond better to probabilities than those of the unseasoned. If this were so, then we could say that the seasoned investors are more emotionally mature (in this limited area)

than the unseasoned; their emotions are more “rational.” But it also seems possible that the seasoned investor has a different, and more rational, view of the risk: on any given day, or in any given option period, he expects, without a great deal of either anxiety or hope, that there will be losses or gains (or both). To him the really interesting probabilities of gain and loss are longer-term, ranging over yearlong or several-year periods of investing. In this way too, his elation or anxiety transcends and fails to track the short-term probabilities (though his investing behavior does track them). But now this itself is a sign of maturity; it is not just that his anxiety fails to track probabilities, but that he is relatively unanxious about them.

6.3 *Irrational Stimuli*

Emotions are sometimes thought to be a bad basis for judgments because of the “irrationality” of the factors that elicit or otherwise condition them. Some such factors are rhetoric and language; evolutionary preparedness; past history of fear conditioning; gender and race; and worldview. Let us briefly consider each of these factors in light of the normative question of how, and how much, emotions should be allowed to influence our judgments of risk.

6.3.1 *Rhetoric and Language*

Two thousand three hundred years ago Aristotle offered, in his treatise *The Art of Rhetoric* (Aristotle 1926, Book 2, Chapters 1–11) a rather sophisticated account of a dozen or so emotion-types and of ways in which a speaker can manage such emotions in an audience. We all know from common experience how much more emotionally powerful narratives are than abstract, analytical discourse, and how the vocabulary and emphasis of a piece of discourse can influence its emotional impact. Modern empirical research confirms that information given in one set of words can be more engaging emotionally than the “same” information conveyed in other words. The importance of imagery – both mental and rhetorical – in eliciting emotions seems to confirm my suggestion that emotions are perception-like states. Miller et al. (1987) found that people who received training in forming vivid images experienced increased arousal by personalized scripts written to elicit anger and fear. Nisbett and Ross (1980) presented people with two descriptions of a death. They found that “Jack sustained fatal injuries in an auto accident” produced less emotion than “Jack was killed by a semi trailer that rolled over on his car and crushed his skull.”

One might say that judgments made under the influence of vivid mental imagery and/or the narrative, concrete, and evocative language that tends to produce more of it, are less rational than judgments made in response to “cool,” abstract, general, and (say) statistical discourse. But to say so would, I think, be to use a tendentious and narrow concept of rationality (see Dickens’s (1854) *Hard Times* for a colorful and devastating critique of such a concept of rationality). If emotions provide a better

way of grasping evaluative truths, as I have argued – a more complete perception and understanding of them – then rhetorical elements in the discourse about such truths may be essential to their practical truth-value, their ability to get the truth *across*. “Cold” language may actually be a way of obscuring truth.

Correlatively, the ability to respond to such language, and even to compensate, by way of imagination, for the rhetorical deficiencies of some presentations of fact, may be important ingredients in intellectual virtue. Some people seem to have, as a natural endowment, better powers of imagination than others. But the study by Miller et al. suggests that the capacity to have vivid mental images can be enhanced by education. (This might be accomplished, in part, by the study of great prose and poetic literature, and the enhancement of one’s intellectual powers by this means would justify such study in a general curriculum designed to promote personal intellectual excellence.) Rhetorical powers (the ability to *produce* evocative thought and language) and a broad emotional receptivity (the ability to *respond* to such language) might be dual and correlated aims of a genuinely liberal education.

We have seen that the strong immediacy of perception involved in emotions can not only enhance the grasp of truth, but can also obscure it and make falsehood compelling to those whose perspective is deficient. Rhetoric is in the interest of this strong immediacy. Those who wield it must do so carefully and responsibly, and those who receive it must be discerning. So once again, intellectual and emotional virtue involves not only emotional receptivity, but also the ability to stand back from emotion, so as to evaluate and manage it.

6.3.2 Past History of Fear Conditioning

As a result of differing personal histories, people differ from one another in the intensity and object-range of their fears. Such a history is likely to have left one with fear-dispositions that are at least partly irrational. In varying degrees we all suffer thus, in our life of judgment, from the ravages of our past, as we profit from its healthy instruction. But that past has also endowed most of us with an ability to judge, dissociate from, and control our fears, and part of the burden of being a morally and intellectually responsible person is to exercise those powers diligently.

6.3.3 Evolutionary Preparedness

We come pre-programmed to fear – or easily to learn to fear – certain kinds of objects, such as snakes, large spiders, and precipices. In the situations of life, the judgments that would be produced automatically by such fears sometimes have to be stopped, in the interest either of truth or of some practice. Practices, with their associated judgments, that run carefully contrary to these fears, such as snake-handling, roofing, and mountain climbing, tend to weaken the fear-reaction and thus to bring it into closer conformity with the truth about the attendant dangers. We learn, for example, to fear certain *kinds* of spiders and snakes, and to have a merely healthy respect for the rest of their tribes. Similarly, we are “naturally” *unafraid* of certain

real dangers, such as infections, poisoning by tasteless chemicals, and moral and intellectual vice, and must learn to fear them.

6.4 Socio-Political Factors

Flynn et al. (1994) found that about 30% of the White Male American population judge risks from various hazards such as cigarette smoking, blood transfusions, ozone depletion, medical X-rays, auto accidents, and nuclear power plants to be quite a bit less than the rest of the American population. Slovic speculates:

Perhaps [certain] White males see less risk in the world because they create, manage, control, and benefit from many of the major technologies and activities. Perhaps women and non-White men see the world as more dangerous because in many ways they are more vulnerable, because they benefit less from many of its technologies and institutions, and because they have less power and control over what happens in their communities and their lives (Slovic 1999, p. 693).

Slovic is saying that these White males may be roughly right in their judgments of these risks insofar as they are risks *to themselves*. Many of the hazards in the list *are* less hazardous to persons of privilege and power than to others. But the judgments they were asked to make were about levels of risk *to society*. If this is the case, then the White male population suffers epistemically from a tendency to confuse risk-levels to others with risk-levels to themselves. This is an epistemically vicious subjectivism, and seems to stem from a kind of egoism or narcissism whose symptom is an inability or lack of disposition to empathize with others, to put oneself imaginatively in their place, to take others seriously into consideration. Egoism is usually thought to be a moral failing, but here we see how it can be an intellectual failing as well, a disability to make accurate judgments. It seems also to be a broadly emotional failing, insofar as it seems to turn on not *caring* very much about risks, as long as they are not risks to oneself. Again, Slovic's data do not imply that emotional perception of risk is not correct, but rather that getting one's risk-judgments right depends significantly on a correct emotional formation.

If emotions are concern-based construals, then moral emotions are construals that are based on moral concerns. This way of seeing the relation between emotions and morality welcomes a much larger range of emotions into the moral club than is usual for philosophers. Usually, the moral emotions are anger and guilt (Rawls 1972; Gibbard 1990/2003): anger at those who transgress against oneself and one's own, and guilt about one's own transgressions against others. But if moral emotions are any emotions based on moral concerns, then joy, gratitude, relief, hope, shame, fear, grief, sadness, compassion, and revulsion, among others, may be moral. And if emotions have the importance for judgments that I have claimed in this paper, then any of these emotions, if they are genuinely and properly moral, can be the basis for the insightfulness, truth, justification, and understanding of moral judgments.

7 Conclusion: The Epistemic Potential of Moral Emotions About Risk

Roeser (2006, p. 696) comments that if someone knows that a rich capitalist is violating her rights by building a dangerous chemical factory in her neighborhood because she is poor and can't defend herself, but feels no anger about this, something is wrong with her judgment. By the same token, we would think that if the rich capitalist knows that what he is doing is wrong, but feels no guilt about it, something is wrong with *his* judgment as well. In this paper I have tried to show how and why this is so. The problem with both judgments is not that they are untrue or unjustified, but that the subjects lack a certain essential kind of understanding in their judgments, a certain insight into their truth. In the case of anger, the emotion supplies a strong immediate *impression* of the injustice in its badness in connection with the badness of the perpetrator; in the case of guilt, the immediate impression is the same, except that the perpetrator is oneself. I am not saying that a person must feel the emotion each time he makes the judgment, to qualify as having this understanding and insight; understanding and insight can be dispositional, so he needs only to be *able* to feel it, and to feel it *sometimes*.

Let me end this paper by pointing out that the other moral emotions can play a similar role. Thus if the neighborhood and its friends have battled the construction of the dangerous factory, and win the battle, those who have worked for this end will feel joy, thus having a strong impression of the good that has been wrought; or relief, thus having a strong impression of the end of their worry and toil; or gratitude for the help they have received, thus having a strong impression of their indebtedness for this benefit and the gracious goodness of the benefactor; of hope, thus having a strong impression of the better future of the neighborhood. If the battle turns out badly, then the moral emotions, in addition to anger, will be such ones as disappointment, regret, and grief. These too will be moral emotions, because they spring from and express a moral concern; and they will present in their subjects' experience the darker aspects of the situation. These emotions other than ones belonging in the fear-family are only incidentally, or indirectly, about risk.

References

- Alhakami, A. S., and P., Slovic. 1994. A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis 14*: 1085–1096.
- Aristotle . 1980. *Nicomachean Ethics*. Translated by W. D. Ross and revised by J. L. Akrill and J. O. Urmsen. Oxford: Oxford University Press.
- Aristotle . 1926. *The "Art" of Rhetoric*. Translated by J. H. Freese Cambridge, MA: Harvard University Press.
- Damasio, A.. 1994. *Descartes' Error*. New York: Putnam.
- Dickens, C.. 1854, 1995. *Hard Times*. Harmondsworth: Penguin Books.
- Finucane, M., A., Alhakami, P., Slovic, and S. M., Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making 13*: 1–17.
- Flynn, J., P., Slovic, and C. K., Mertz. 1994. Gender, race, and perception of environmental health risks. *Risk Analysis 14*: 1101–1108.

- Gibbard, A.. 1990/2003. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- Gladwell, M.. 2005. *Blink: The Power of Thinking Without Thinking*. New York: Little, Brown, and Company.
- Loewenstein, G. F., E. U., Weber, C. K., Hsee, and N., Welch. 2001. Risk as feelings. *Psychological Bulletin* 127: 267–286.
- Miller, G. A., D., Levin, M., Kozak, E., Cook, A., McLean, and P., Lang. 1987. Individual differences in imagery and the psychophysiology of emotion. *Cognition and Emotion* 1: 367–390.
- Nisbett, R., and L., Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Rawls, J.. 1972. *A Theory of Justice*. Oxford: Oxford University Press.
- Roberts, R. C.. 2003. *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.
- Roberts, R. C., and W. J., Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Clarendon Press.
- Roeser, S.. 2006. The role of emotions in judging the moral acceptability of risks. *Safety Science* 44: 689–700.
- Slovic, P.. 1999. Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis* 19: 689–701.
- Weber, E. U., and C. K., Hsee. 1998. Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards risk. *Management Science* 44: 1205–1217.

The Moral Risks of Risky Technologies

Peter Goldie

1 Introduction

In his *Civilization and its Discontents*, Freud, writing in 1930, noted our increasing dependence on technologies – ships, aircraft, spectacles, telescopes, cameras, gramophones, telephones, and so on. He said, “Man has, as it were, become a kind of prosthetic God. When he puts on all his auxiliary organs he is truly magnificent; but those organs have not grown on to him and they still give him much trouble at times. Nevertheless, he is entitled to console himself with the thought that development will not come to an end precisely with the year 1930 A.D. Future ages will bring with them new and probably unimaginably great advances in this field of civilization and will increase man’s likeness to God still more. But in the interests of our investigations, we will not forget that present-day man does not feel happy in his Godlike character” (Freud 1930/1985, pp. 279–280).

Since 1930, there have indeed been “unimaginably great advances” in technologies: computers, satellites, GPS navigation systems, mobile telephones, robots, embodied conversational agents (ECAs), avatars, androids, and so on. But our auxiliary organs still “give us much trouble at times”: they go wrong; they take on a life of their own; and they are often incomprehensible in their function.

So in spite of these advances we still do not feel entirely happy in our “Godlike character”. Quite what the unhappiness consists of is not entirely clear. This is an empirical issue on which I do not wish to reach any definitive conclusions. My main interest here will be in the normative issues. However, it would seem that, to a considerable extent, we have ambivalent or mixed feelings towards our auxiliary organs – and these are manifested in particular in our emotional responses towards them, and in the other ways in which we interact with them. From now on I will limit my discussion of technologies to computers, robots, avatars, ECAs, and other

P. Goldie (✉)

Department of Philosophy, University of Manchester, Manchester, UK
e-mail: peter.goldie@manchester.ac.uk

kinds of emotion-oriented technologies (EOTs), many of which are known as semi-intelligent information filters (SIIFs); in general, these are kinds of technologies with which we tend to *interact* (often emotionally), and not just act towards, as we do, for example, towards those which Freud discusses. It is largely for this reason that the interesting issues which I want to discuss here arise.

2 Ambivalence in Our Behaviour Towards Technologies

On the one hand, it seems that we relate to these technologies as if they are simply what they in fact are: inanimate objects, incapable of any kind of thought or feeling, and thus no more deserving of any kind of *human* interaction as might be a screw-driver or a cabbage. And yet even here, when they “give us trouble”, we verbally and physically abuse them in ways that are somehow oddly expressive of our frustration (de Angeli et al. 2006). For example, we complain that the thing has a “mind of its own”, we shout and swear at it (“Come on, you damned thing, work!”), and we beat it, often to our own detriment as well as to the machine itself. These kinds of actions are clearly expressive of emotions such as frustration and anger (Hursthouse 1991), but the manner of expression is in many ways peculiar to this kind of object: we are, for example, less likely to shout at a tube of toothpaste if it fails to work, whereas it is typical behaviour towards a malfunctioning computer, or to the pre-recorded telephone message from the airline company, telling us they are “sorry” to keep us waiting, and that our call is “valuable” to them.

Ambivalence is revealed in other empirical findings that we are also (and somewhat contradictorily to our aggressive behaviour) capable of behaving politely towards computers, unconsciously treating them as we might humans, whilst at the same time denying that we think of them as human (Nass and Reeves 1996).

Yet further evidence of ambivalence is found in studies which have involved carrying out Milgram-style experiments on what are known by participants to be inanimate avatars and robots (Milgram 1974; Slater et al. 2006; Rosalia et al. 2005; Bartneck et al. 2006; Bartneck and Hu 2008). In one series of experiments on an avatar, a female virtual person, the investigators concluded as follows: “Our results show that in spite of the fact that all participants knew for sure that neither the stranger [the avatar] nor the shocks were real, the participants who saw and heard her tended to respond to the situation at the subjective, behavioural and physiological levels as if it were real” (Slater et al. 2006, p. 1).

And finally, of course, there is the vexed question of what to make of the “uncanny valley”, as introduced by Masahiro Mori (Mori 1970), and now much discussed in robotics and computer science. What Mori argued was that our emotional attitudes towards robots change as the robots become more and more similar to human beings (in behaviour, in facial and verbal expression, and so on). We are thus more comfortable with a humanoid robot than an industrial robot, and yet when the robot becomes even closer in appearance to a healthy human but is still clearly *not* human, our feelings of comfort and familiarity decline: we are in the uncanny valley. There are a number of explanations that have been put forward for this kind

of reaction that we have: that the robots are “bukimi” in Mori’s sense – weird, ominous, eery; that they give rise to disgust; that they deviate from the norms of physical beauty; that they frustrate our (largely unconscious) expectations; that they give rise to fear of death (MacDorman and Ishiguro 2006).

These emotional responses and patterns of behaviour, expressive of our ambivalence, are generally not of the kind that can be seen as rational, in the way that, for example, fear of a savage dog would be rational. They are, rather, more visceral, more primitive.

3 Not Ambivalence in Belief

It is important to appreciate here that this ambivalence in our emotional responses and behaviour does not seem in any way to be grounded in ambivalence in our *beliefs* about whether or not computers, robots, and so on are minded, and thus capable of thoughts and feelings. The point can be put in terms of the more general contrast between two kinds of consciousness: what the philosopher Ned Block (1997) has called *access consciousness*, as contrasted with *phenomenal consciousness*. Roughly, access consciousness is the kind of consciousness involved in mere cognition – information storage and processing for example. So, for example, the capacity of something to recognise a threat and to respond with evasive behaviour has access consciousness. And, still as part of access consciousness, a more complex organism might also be capable of recognising its own internal states, such as the state which represents *that* it is threatened and *that* a certain kind of evasive response is called for. Phenomenal consciousness, in contrast, is what is involved when *there is something that it is like* for the organism – in this case, where there is something that it is like to feel fear (Nagel 1974). There is something that it is like to be a human, a dog, or a cow – they all have phenomenal consciousness, and they all can experience fear – but there is nothing it is like to be a computer, or a robot.

Could there ever be something that it is like to be a robot – could a robot ever, for example, experience fear? As science fiction literature and film attest, we feel unsettled by the apparent fact that, in the fiction, these non-animal things are capable of emotional feelings, and we feel inclined to empathise with them in this respect. Consider, for example, the Nexus-6 replicants in *Blade Runner* (Ridley Scott 1982) who are programmed with a fail-safe device to cease functioning after 4 years in case they start to develop empathy; and the computer Hal in *2001: A Space Odyssey* (Stanley Kubrick 1968), which seems to be motivated emotionally, by revenge or envy perhaps, and seems to suffer as his systems are shut down. But these are thought experiments, and there is no evidence that adults are inclined to believe that actual technologies are capable of experiencing emotions (Picard 2002). The ambivalence in our behaviour and emotional responses, and even in our empathetic responses on some occasions, does not then seem to be grounded in an ambivalence or uncertainty in belief. And this stands in marked contrast to how we might, for

example, be ambivalent or uncertain in our beliefs about what is going on in the struggling trout on the end of the fishing line, or in the harpooned whale in its final death throes; in such cases, we might indeed be unsure of what is going on in the living creature.

So far, then, the discussion has been restricted to the empirical question of what our attitudes and behaviour are towards technologies of the kinds I have been focusing on, and my supposition is that these involve ambivalence of emotion and behaviour, but not ambivalence or uncertainty of belief.

Be that as it may, it is to the normative question that I now want to turn, and this will be the focus for the remainder of this chapter. What sort of attitudes and behaviour *ought* we to adopt towards these technologies?

4 The Rationality of Our Responses to Technologies

One obvious thought might be suggested to begin with: whatever else our attitudes and behaviour ought to be, they ought at least to be rational. However, on examination this thought runs the risk of proving either too much or too little. The point can be made by reference to a parallel argument in relation to our emotional engagement with fictional characters, an argument which is supposed to reveal a paradox of irrationality – the so-called paradox of fiction. It is paradoxical that each of these three propositions is intuitively acceptable: that we feel emotions towards fictional characters; that to be rational in feeling an emotion towards something we must believe that thing to exist; and yet we do not believe that fictional characters exist. Colin Radford, for example, has argued extensively that there is no acceptable reply to this paradox, and that it shows that our emotional responses to fictional characters are irrational: inconsistent and so incoherent (Radford 2001).

This conclusion, if true, would surely prove too much if it showed that we ought not to have emotional responses to fictional characters, simply on the grounds that such responses are irrational. And it would prove too little if it showed only that we have manifested a form of irrationality, without any implication that it ought not to be encouraged in other respects. In my view – which I cannot argue for here – the central difficulty with the so-called paradox of fiction is that the notion of rationality that is at work in setting up the paradox is so thin (Goldie 2009) that it has little force in recommending how we ought, all things considered, to think and feel.

It can be readily seen how a similar paradox could be set up for our emotional responses to, for example, the “cruel” treatment of an avatar of the kind found in the Milgram-style experiments that I mentioned above. The paradox would go something like this: we respond (let us assume) with moral concern to the treatment of the robot; we ought rationally to respond with moral concern to the treatment of something only if we believe that thing to have thoughts and feelings; and yet we do not believe that robots have thoughts and feelings. This argument might indeed show that this kind of response is irrational, but still, as with the parallel

argument about our emotional engagement with fictional characters, it either shows too much or too little. What we need to do is to consider the wider normative considerations, both moral and practical, that enter into addressing the question of how we ought, all things considered, to relate to technologies of the kinds I am concerned with.

5 Instrumental and Non-instrumental Value

Some things have merely instrumental value: something which is of instrumental value is to be valued only in so far as it is good of its kind, so that it performs its function well. For example, a knife is instrumentally valuable only in so far as it is able to cut; once it ceases to be able to perform that function, it ceases to be of value.

Shocking as it might be to us, Aristotle thought that slaves were valuable only in this way: “The slave”, he said, “is a living tool” (*Nicomachean Ethics*, 1161 b 4). But we should not, in recoil from this, turn to rejecting the idea that people should ever be thought of as having instrumental value. For it is undeniable that the taxi-driver, the housekeeper, the nanny, the man in the ticket office, can all have this kind of value. Rather, we should accept that humans can have instrumental value, but we should at the same time insist that they also have non-instrumental value, that they are of value for themselves, and not only for some further purpose. This is what is behind the “merely” in Kant’s famous claim: “So act that you always treat humanity . . . always at the same time as an end, never merely as a means” (1785/1964, p. 429).

It is controversial quite what is involved in treating people as ends, but I do not need to appeal to anything more here than a negative duty which is at least part of what is involved: the duty not to abuse people, not to treat them cruelly or aggressively. Of course more than that is involved in how we ought to treat people, but this will not be my concern here, for reasons which will emerge.

There is no doubt that technologies have instrumental value – when they work. The question that is pressing is whether, like people, they also have non-instrumental value, and if so, of what kind. There is, in fact, a range of possible sources of non-instrumental value here: we do not have to attribute non-instrumental value to technologies for the same reasons – essentially moral reasons – as we have to attribute this kind of value to people. I will briefly consider three other possible sources of non-instrumental value before turning to moral reasons of the kind that Kant had in mind.

One possible source of non-instrumental value that might apply to something such as a tool or a piece of technology is sentimental value (Hatzimoyisis 2003). For example, if I have a fountain pen that was given to me by someone I hold very dear, then I might well continue to treasure that pen even after it has ceased to perform its function well – even after it no longer works. There is no doubt that technological things sometimes do have sentimental value in this way: for example, some people

hang on to old and highly unreliable laptops just because they now have this kind of value for them. But it should be noticed about this kind of value that the value depends on the existence of the relevant associations, and it follows that the value is agent-relative in the sense that something which is of sentimental value for me need not be of sentimental value for you, just because it does not possess the relevant associations for you.

A second possible source of non-instrumental value of technologies is that one comes to consider them to be, in some sense, friends or companions. (For discussion of the value of friendship, see Stocker (1976).) Again, there are no doubt instances of this to be found, such as the way children behave towards their Tagamochi toys. But this value, like sentimental value, is agent-relative, and, moreover, there are perhaps concerns to do with the possibility of psychic disharmony that might undermine this kind of attitude. (Note here that I make the point not in terms of irrationality, but in wider terms to do with possible damage to the individual.)

Thirdly, there is aesthetic value, which, unlike sentimental value and value as friends or companions, is not agent-relative. A distinction of Kant's here is helpful in distinguishing two ways in which a piece of technology might have aesthetic value. Kant, in his great work on aesthetics, *The Critique of Judgement* (1790/1953), distinguished between free and dependent beauty. As Kant put it, "The first presupposes no concept of what the object should be; the second does presuppose such a concept and, with it, an answering perfection of the object" (Kant 1790/1953, p. 72). Interpretation is famously tricky here, but the essential idea is that something is freely beautiful if we can judge it to be beautiful without having a clear idea of what kind of thing it is or what its purpose is; Kant's example was the beauty of a flower. In contrast, something is dependently beautiful if we need to know what kind of thing it is, and what its purpose is, before we can judge its beauty; as Kant says, it is "ascribed to Objects which come under the concept of a particular end" (Kant 1790/1953, p. 72). For example, we might need to know that something is a rapier, and what the purpose of a rapier is, in order to judge its beauty: our judgement depends on this prior knowledge. Kant's own examples included men, horses, and buildings (Scarre 1981).

It strikes me that pieces of technology are capable of possessing either or both of these kinds of aesthetic value. The enormous NASA computer facility containing cabinet after cabinet of quietly humming mainframes might possess dependent beauty, because we need to know that the purpose of this facility is to track the movement of the stars in the Solar System if we are to appreciate its beauty. In contrast, perhaps the latest Apple laptop is freely beautiful: its design is such that we can admire its beauty without first needing to know what it is or what is its purpose.

So there are these three kinds of reasons for attributing non-instrumental value to technologies. Each of them is, I think, interesting in its own right, and may well have application in particular cases, but what I am seeking is a kind of reason that is somewhat more universal in its application than these, and with that in mind I now turn to moral considerations.

6 Moral Reasons for Valuing Technologies

Why might we think that technologies have moral value of a kind which is non-instrumental, so that they are valuable not only for some further purpose? Again, there are a number of possibilities here, and I want to eliminate some before turning to what I think is the most important moral consideration.

First, we might think that technological items such as robots have rights. Peter Singer has argued for a number of years that non-human animals have rights (for example, in Singer 1977), and it has even been suggested recently (in a report titled “Robo-Rights” commissioned by the UK Office of Science and Innovation’s Horizon Scanning Centre in December 2006) that rights could indeed be extended to robots. But even if we reject that idea as sheer madness (and the report was highly criticized at the time), we might still think that we have duties towards them. More interesting, though, is the thought that we have duties *with regard to* them, and it is this thought that I will turn to later.

Secondly, we might think that we should attribute moral value to robots and so on because they are sentient, or at least because we are not certain whether or not they are sentient, and we should, so to speak, give them the benefit of the doubt. But this is something that I considered earlier. We do not believe that they are sentient, and we do not seem even to believe there to be any doubt about the matter, so no moral choice arises here, as it might with fish or whales for example (Dennett 1996), even if we do sometimes empathise with them as if they are sentient. And it seems to me that we are right about this. Leaving aside any science-fiction future possibilities, we are in fact right to believe the contrary: to believe that current technologies do not possess phenomenal consciousness.

Even so, perhaps we should attribute moral value to them at least on the grounds that they do seem to possess *intelligence*, in the sense that they seem to possess *access* consciousness (Bartneck et al. 2006). Intelligence could be something that we should value in the world not only for its instrumental value. Perhaps, but I will leave that interesting thought, like the others, in suspense in order to turn to the moral considerations that I think bear most weight here, and has the widest range of application to technologies beyond just robots, computers, and other technologies that seem to possess intelligence.

Here is the central idea. The way we treat technologies can be expressive of our personality. Consider, for example, the person who regularly shouts at his computer for not working as he wants, bashing the “Enter” key in irritation and frustration. What might begin as behaviour towards just this computer can easily become more general and expressive of personality traits, such as irritability and short-temperedness, directed towards a wide range of technologies: towards the computer, towards the ticket machine in the railway station, towards the airline’s automatic telephone answering system, and so on. This irritable and short-tempered behaviour can then easily become generalised beyond technologies to people as well: towards the person in the ticket office as well as towards the ticket machine; towards the airline official on the telephone as well as towards the automatic answering system.

These officials come no longer to be treated with the respect that should be accorded to them as persons, coming to be treated merely as means and not also as ends in themselves. I think we all know the type who behaves like this: the kind of person who sees everyone else as existing only to help him achieve his goals, never accepting that others might have goals of their own.

Personality traits of this kind are largely a matter of habit. To begin with we become habituated to treating our technologies in this way, and this then readily extends to the treatment of people whom we use as means. Ultimately, if it becomes endemic in the population, we often find that it results in a dystopia, where a whole class of people are treated merely as technologies: the workers in Fritz Lang's film *Metropolis* (1927), or in Chaplin's *Modern Times* (1936). The central idea, then, is that there is a kind of slippery slope here, from the way we treat technologies, to the way we treat people. Largely as a matter of habit, we move readily from treating technologies merely as means to treating people merely as means. And we should cultivate our personality traits to make sure that we do not slide down this slippery slope, and, in order to do this, we should avoid abusing technologies. Thus we would be wrong to think that abusing technologies, in the privacy of one's own home or workplace, is a harmless activity.

There is an analogy here with Kant's discussion of our duties with regard to non-human animals. (In what follows I am much indebted to the discussion in (Korsgaard 2004).) Kant's idea was that we tend to mistake our duties with regard to non-human animals for a duty *towards* those animals – a duty that that we have in virtue of those animals having some kind of call on us. (Kant called this an “amphiboly”.) Kant thought that the only kind of thing that we have duties towards is human beings (ourselves and others) as rational animals. He maintained, in contrast, that we have duties *with regard to* other animals; the mistake (the amphiboly), he thought, was to think that we have duties *towards* them. Non-human animals, Kant thought, are “analogues” of humanity, and our duty is not to them, but to ourselves, “to cultivate our duties to humanity” by acting and feeling dutifully *in respect of* non-human animals. Kant put it thus:

With regard to the animate but non-rational part of creation, violent and cruel treatment of animals is . . . intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls his shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relation with other people. The human being is authorized to kill animals quickly (without pain) and to put them to work that does not strain them beyond their capacities (such work as he himself must submit to). But agonizing physical experiments for the sake of mere speculation, when the end could also be achieved without these, are to be abhorred. – Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs *indirectly* to a human being's duty *with regard to* these animals; considered as a *direct* duty, however, it is always only a duty of the human being *to himself* (Kant 1797/1996, p. 443), cited in part in Korsgaard (2004, pp. 90–91).

Now, I do not want to consider whether or not Kant's views about non-human animals is correct, or whether an alternative view (such as that of Peter Singer) is to be preferred, a view that ascribes rights to non-human animals, so that we have consequent duties *towards* them and not merely *with regard to* them. For we can reject

Kant's views about non-human animals, but still insist on the correctness of the parallel view in relation to technologies. So technologies have no rights and we have no consequent duties *towards* them. But we have duties *in respect of* technologies. This is the duty to ourselves to cultivate our personality traits in respect of them, because acting in accordance with this duty cultivates our acting dutifully towards people, whom we should always treat as ends.

Recall here, though, that I am merely arguing that this duty with regard to technologies extends only to not treating them badly or abusing them in the various ways I have been discussing. It does not extend to the kinds of positive duties that are involved in respect for people – nor, indeed, to the gratitude for long service that we accord to the horse or the dog! So the range of personality traits to focus on will include, for example, curtailing irritability and short-temperedness, and not on, for example, developing gratitude and generosity.

It might be complained that what I am proposing is motivationally paradoxical, in the sense that I am advocating that we should be motivated to treat technologies as if they have non-instrumental value in spite of knowing that they do not have such a value, and that we should do so in order to avoid a slide down the slippery slope. The paradox, according to the complaint, is that the motivating reasons for adopting the practice are in fact external to the practice whilst we are supposed to treat them as if they are internal – as if technologies really do have non-instrumental value so that our duties are towards them. But the motivational paradox is not as tight as the complaint suggests. Consider, for example, how one might begin jogging in the morning in order to lose weight, but one appreciates that in order to do this every morning one must enjoy running for its own sake. It sounds paradoxical to say “I should enjoy running for its own sake in order to lose weight”, but the motivational pattern is clear enough. Many of our motivations for practicing certain kinds of behaviour begin as external, but in the knowledge that the best way of keeping up the practice is for the motivations to become internal to the practice.

A further complaint against what I am suggesting is that my claim rests on the idea that there really is a slippery slope here, and this is open to question. Indeed, there are some slippery slope arguments that are problematic, but this is not one such. In his paper “What slopes are slippery?”, Bernard Williams made the distinction between two types of slippery slope argument: the “arbitrary result” argument; and the “horrible result” argument. The latter relies both on the argument that there is “no point at which one can non-arbitrarily get off the slope once one has got on to it”, and on the further argument “that there is a clearly objectionable practice to which the slope leads” (Williams 1995, p. 213). As an example of the first, Williams considers the claim that the extension of some kind of married person's tax relief, from couples who are legally married to some other couples, would put one on a slippery slope where any cut off point in the relief would end up as arbitrary. As an example of the second, Williams mentions the argument against *in vitro* fertilization of human ova.

My argument is of the second kind: the “horrible result” is the failure to treat other people as they ought to be treated: not merely as means but also as ends in

themselves. And we come to do this, so the argument goes, as a consequence of abusing technologies in various ways. So the argument against abusing technologies is, in this sense, consequentialist. From this it will be evident that it is necessary for this argument to go through that there be a plausible *psychological* slippery slope, from the abuse of technologies to the abuse of people, as, for example, there is a psychological slippery slope for the alcoholic in moving from one drink to one drink too many (Williams 1995, p. 218). In respect of my argument, the psychological slippery slope involves interesting issues concerning the relation between personality traits, moods, and emotions (Goldie 2000; Goldie 2004). Consider the person who starts the day before going to work abusing his computer, his mobile phone, and various other technologies. His emotion towards these things is one of anger – anger that they will not work and interact with him as they ought. These emotions put him in an irritable mood – one where he is prone to get angry at other things that will not do as he wants: towards his children for not eating their breakfast; and then later in the morning towards the man in the ticket office for not dealing with his request as he thinks appropriate. And these emotions and moods, over time, consolidate into a personality trait: into the disposition to get angry and to abuse people in general as well as technologies in general. The psychological slippery slope, then, does not involve the risk that there is “some motive . . . to move from one step to the next” (Williams 1995, p. 218). The risk, rather, is that one becomes habituated in feeling and behaving a certain way, and thus one unthinkingly moves, out of habituation grounded ultimately in a personality trait, from one step to the next.

Finally, there might be a concern that my idea, that we should behave with respect towards technologies by not abusing them, runs the risk of putting us on a different slippery slope: this time from treating technologies with respect by not abusing them, to treating them with respect by treating them just as we would treat human beings. It might be said, in support of this concern, that a similar slippery slope can arise in our treatment of animals: the animal lover sometimes comes to treat non-human animals and humanity with *equal* respect, or even, at the extreme, with *more* respect than humans, as perhaps is sometimes found with animal rights campaigners, who abuse, terrorise, or even kill their fellow human beings in order to protect other animals. Treating animals with respect ought not to turn into treating them just as we treat humans. And, of course, the same point applies to technologies – a fortiori one might reasonably think. I think we can accept that there are some grounds for this concern with small children, as evidenced by the tyranny that Tagamochi toys can have over their lives. But this particular slippery slope argument fails, because there is not a genuine *psychological* slippery slope here. Negative, abusive behaviour of the kind I have been concerned with is habitual and characteristically not reason-based, so that one can all too easily slide from abusing technology, to abusing non-human animals, and then to abusing people. In contrast, positive, caring behaviour is characteristically reason-based and not habitual, so there is no reason to think that this slippery slope is a concern for most adults. Just as most of us are able to distinguish between our positive duties with regard to

non-human animals from our duties towards humans, I think we can do the same with technologies. Things are different, though, with our bad habits.

7 Conclusion

Freud's remarks in 1930 show us that, in a sense, there is nothing new in our relation to technologies: in spite of the advances, they continue to give us difficulties at times. And yet, in another sense, there *is* something new. Today, we *interact* with many technologies in ways that we did not in Freud's day: we interact with robots, with avatars, with androids, with EOTs, and with ECAs. Like non-human animals, but in a different way, they have become, to use Kant's term, *analogues* of humanity. And, because of this, there is now a particularly slippery psychological slope from the abusive ways in which we can treat these technologies to the abusive ways in which we come to treat humanity. This slope is to be avoided, and the way to do so is to cultivate our personality traits so that we treat technologies with respect, to the extent of not abusing or otherwise behaving badly towards them.

Finally, I should say something about the relationship between the normative and the empirical issues in this chapter. I said that I would focus mainly on normative questions rather than empirical ones, but in the end it is important to accept that my slippery slope argument depends on certain facts about human psychology, and it is, in just that sense, empirical.

Acknowledgments My thanks to Sabine Roeser for her support as editor of this volume, and to Roddy Cowie and others for their contribution to, and discussion of, work on which this chapter is based, and to HUMAINE (Human-Machine Interaction Network on Emotion), a Network of Excellence in the EU's Sixth Framework Programme (Contract no. 507422).

References

- Aristotle. *Nicomachean ethics*, Translated by W. D. Ross and revised by J. O. Urmson. In *The Complete Works of Aristotle: The Revised Oxford Translation, Vol. 2*. J. Barnes, ed., Princeton: Princeton University Press.
- Bartneck, C. et al. 2006. To kill a robot. *Proceedings of the Workshop on Misuse and Abuse of Interactive Technologies in cooperation with the Conference on Human Factors in Computing Systems CHI2006* Montreal, Canada.
- Bartneck, C., and J., Hu. 2008. Exploring the abuse of robots. *Interaction Studies – Social Behaviour and Communication in Biological and Artificial Systems* 9: 415–433.
- Block, N. 1997. Biology versus computation in the study of consciousness. *Behavioral and Brain Sciences* 20: 159–165.
- de Angeli, A., S., Brahmam, P., Wallis, and A., Dix (2006). Misuse and abuse of interactive technologies. *Proceedings of the Workshop on Misuse and Abuse of Interactive Technologies in cooperation with the Conference on Human Factors in Computing Systems CHI2006* Montreal, Canada.
- Dennett, D. 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- Freud, S. 1930/1985. Civilization and its discontents. In *The Penguin Freud Library, Vol. 12: Civilization, Society and Religion*. S. Freud, ed., 243–340, London: Penguin.

- Goldie, P. 2009. Thick concepts and emotion. In *Reading Bernard Williams*. D. Callcut, ed., 94–109, London: Routledge.
- Goldie, P. 2004. *On Personality*. London: Routledge.
- Goldie, P. 2000. *The Emotions: A Philosophical Exploration*. Oxford: Clarendon Press.
- Hatzimoyisis, A. 2003. Sentimental value. *Philosophical Quarterly* 53: 373–379.
- Hursthouse, R. 1991. Arational actions. *The Journal of Philosophy* 88: 57–68.
- Kant, I. 1785/1964. *Groundwork of the Metaphysics of Morals*. Translated by H. J. Paton, New York: Harper & Row.
- Kant, I. 1790/1953. *The Critique of Judgement*. Translated by J. C. Meredith, Oxford: Oxford University Press.
- Kant, I. 1797/1996. *The Metaphysics of Morals*. Translated by M. Gregor, Cambridge: Cambridge University Press.
- Korsgaard, C. 2004. Fellow creatures: Kantian ethics and our duties to animals. *Tanner Lectures on Human Values* 24: 79–110.
- MacDorman, K., and H., Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7: 297–337.
- Milgram, S. 1974. *Obedience to Authority: An Experimental View*. New York: Harper and Row.
- Mori, M. 1970. The uncanny valley. *Energy* 7: 33–35.
- Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83: 435–450.
- Nass, C., and B., Reeves. 1996. *The Media Equation: How People Treat Computers*. Cambridge: Cambridge University Press.
- Picard, R. 2002. What does it mean for a computer to ‘have’ emotions? In *Emotions in Humans and Artefacts*. R. Trapp, P. Pettaand, and S. Payr, eds., Cambridge, MA: MIT Press.
- Radford, C. 2001. Paradoxes of emotion and fiction. *The Philosophical Review* 110: 617–620.
- Rosalia, C., R., Menges, I., Deckers, and C., Bartneck (2005). Cruelty towards robots. *Robot Workshop – Designing Robot Applications for Everyday Use*, Göteborg.
- Scarre, G. 1981. Kant on free and dependent beauty. *British Journal of Aesthetics* 21: 351–362.
- Singer, P. 1977. *Animal Liberation: Towards an End to Man’s Inhumanity to Animals*. Boulder, CO: Paladin.
- Slater, M., A., Antley, A., Davison, D., Swapp, C., Guger et al. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* 1: 1.
- Stocker, M. 1976. The schizophrenia of modern ethical theories. *The Journal of Philosophy* 73: 453–466.
- Williams, B. 1995. Which slopes are slippery? In *Making Sense of Humanity and Other Philosophical Papers*. B. Williams, ed., 213–223, Cambridge: Cambridge University Press.

Ethical Imagination: Broadening Laboratory Deliberations

Simone van der Burg

Usually ethicists of technology pass judgment on a technology when it is already developed and ready to be put on the market. But at that point it is often too late to change anything about a technology. As Collingridge (1980) showed in his well-known analysis *The social control of technology*, many parties –such as (public) research funding institutions, researchers, designers and producers – have invested time, effort and money in the development of the technology and when it is ready, they have an interest to put it on the market. At that point it is difficult for ethicists to prevent this from happening.

Collingridge's claim that attempts to change or steer technology often come too late, has been influential: it has led to the engagement of social scientists –and since recently also ethicists – in an earlier phase: that is, during research and development. This early involvement of social scientists or ethicists offers the opportunity not only to try to influence the decision whether or not to implement the technology, but also to co-shape the development of the new technique. However, it has also been noted that it is not realistic to expect that this will lead to *control* over technology. (Rip et al. 1995) The construction of a new technology, as well as the process in which it is brought on to the market, are subjected to the decisions of many people and institutions; such as scientists, research funding institutions, producers and designers.¹ So the social scientist or ethicist who is engaged in the R&D phase will be just one party among others which influences the final shape that the technology will acquire and whether and how it will be sold and used.²

S. van der Burg (✉)

Philosophy Department, Twente University, Enschede, The Netherlands
e-mail: S.vandenburg@utwente.nl

¹For example, the decision of a commercial producer not to invest in the further development of, say, a preventive medical technology for reasons of economic risks, may motivate researchers to search for a public research funding institution, which strongly influences the aspects of the technology that will have the chance to be investigated. And accordingly, the resulting technology will be different.

²These sociological insights have also been relevant to some forms of ethics of technology, in which case-studies take the form of a story about an individual or organization which faces an

This differentiated view of power and decision-making that sociological studies of the R&D phase have produced, is instructive if we want to understand the role that social scientists or ethicists could adopt, who are embedded in the research and development phase. They usually do not seek to “control” one single decision – for what decision should that be? – but take a role in the negotiation that takes place between the parties who influence the shape that the technology will acquire and its implementation in society. They could for example (1) study decision-making processes in the laboratory and mirror them back to the researchers, and contribute in that way to self-awareness and self-reflexivity of the participants (Fisher 2007), (2) invite new parties at the negotiation-table during the research phase such as citizens or patients, and understand the conflicts between their interests to be the relevant ethical issues (Zwart et al. 2006), or they can (3) engage in an imaginative anticipation of the effects that the technology that results from the research will have on the quality of human (social) life, based on conversations about the quality of life with different stakeholders, and bring those views into the conversation with scientific engineers about the future scenarios of the technology their research contributes to. (van der Burg 2009).

This contribution will follow this last approach, which focuses on quality of life issues, and which I developed during my own work as an embedded ethicist. It will offer a case-study, which is based on my embedded ethical work in a scientific engineering context and focuses on research into a medical technology called an acousto-optic monitoring device, which is intended for the non-invasive monitoring of chemical substances in the blood, such as oxygen, glucose and cholesterol. This technology is currently (2009) being researched by a small research group consisting of three people: a PhD student, a technician and a professor. These researchers are part of the Biophysical Engineering Group at the University of Twente (the Netherlands). Research into this acousto-optic monitoring device is still in an early phase: during the past 3 years it took place in the controlled environment of the laboratory. The technology, however, is an original version of a broad family of technologies, which includes optical sensing techniques for oxygen and glucose (See for example: Aoyagi 2003; Sieg et al. 2005) and acousto-optic and photoacoustic techniques for the non-invasive imaging of bowels or tumours in the body (See: Manohar et al. 2007; Xu and Wang 2006; Selb et al. 2001). Some of these technologies are already being used, and some are still in the process of being researched at different locations in the world. Because these technologies are related, they likely have some similar features, and in so far as these features are problematic, the different technologies might share these problems. I want to focus here especially on problems that these technologies likely have when they are used on people with dark skin.

After a brief introduction into this technology and its history in part two, in part three possible future ways will be anticipated in which this technology could affect the quality of human life when it will be used. Here, special attention will be paid to

either-or decision; such as, should this technology be put on the market or not? Or, should I prevent the challenger-launch or not? (Lynch and Kline 2000)

the way in which this technology could affect the emotions of people. Before I start the case-study, however, I want to explain briefly how I am going to approach “the quality of human life”, and how it is connected –and not connected – to the concept of “risk” which is the main topic of this volume.

1 “Risk” and the “Good Life”

The term “risk” includes a broad variety of meanings, but ethicists most commonly take risks to refer to what I will call here “hard impacts”.³ Hard impacts refer to concrete *harms* that a treatment or technology may inflict, which can be counted in numbers of injuries or amount of losses of health or life. Usually ethicists talk about harms for human beings, but sometimes they include also harms for animals or nature. The term “risk” indicates only vaguely that there is a chance that this harm occurs, and sometimes attempts are made to be very precise as to how big a chance that is. In the context of quantitative risk assessment, for example, an attempt is made to be more precise as to how “probable” the risk is; in this literature “risk” is expressed in terms of a number that indicates the average annual probability that a fatality occurs, which is usually based on past experiences of fatalities in the same or a similar context, for example, the probability of the occurrence of a car collision is based on past car accidents in the same area.⁴

There has of course been a lot of debate within ethics about these characterizations of “risk”. Classic questions include, for example, *by whom* risk is to be defined, and whether to assign it subjective or objective meaning (Teuber 1990); or there are authors who analyze the term “probability” and state that uncertain decisions are often treated as if they were decisions under probability (Hansson 1996; 2003). But the critique I am most interested in here aims at the interpretation of “risk” in terms of harms to bodies that can be counted, which are called here “hard impacts”. In this article, I am interested in tracking the possible consequences that the use of acousto-optic monitoring will have for the quality of life, which depends on many more factors than life and health alone. I therefore side with authors who adopt a much broader view of “risk” and also include psychological harms, which refer to impacts on people’s emotion, experience, relations to others and ways to deliberate and act, etc. (Such as Malek and Kopelman 2007) It is these impacts that will be called “soft impacts”.

If “risks” can include hard as well as soft impacts, the term seems to be sufficiently rich to include the kind of impacts that technologies may have on the quality of human life. This interpretation of risks as a combination of hard and soft impacts would go together well with an approach to the “good life” which has been

³With thanks to my colleague Tsjalling Swierstra to whom I owe the distinction between “hard” and “soft” impacts.

⁴In risk-benefit analysis a risk is also expressed in terms of a number, but here the number stands for monetary value assigned to a negative outcome such as an injury or loss of life.

developed by Martha Nussbaum and Amartya Sen, and is known as the “capability approach”. (Nussbaum and Sen 1993) “Capabilities”, according to these authors, are abilities that need to be developed and fostered in order to function well as a human being and reach a state of wellbeing, or of “flourishing”. Examples of capabilities are life, bodily health and bodily integrity, but also the capability of the senses, imagination and thought, emotions, practical reason, affiliation with others, play etc. The aim of this capability-approach is to offer a measure for the quality of human life which is sufficiently rich, so that it can assess the way in which people can conduct their lives in a specific context. On the basis of this list of capabilities, questions can be asked about the life expectancy of people in a specific area, the availability and type of medical services of health care, the quality and availability of education and labour, the types of relations people have with employers, and the political affiliations they are able to engage in, the freedoms they have in conducting personal relations, but also how people in society are able to imagine, to wonder, feel emotions, and play etc.

Nussbaum and Sen do not pay specific attention to the kind of technologies that are available in a certain area. However, the availability of technologies are likely to make possible or help the development of some capabilities, and can also endanger or put obstacles to such development.⁵ Similarly, it is possible to imagine how a *new* technology, which is not yet being used, could do that when it becomes useable. If “risks” are understood in a broad sense, and include hard as well as soft impacts, it seems proper to call an anticipated obstacle that a new technology could impose on human development a “risk”: it refers to a way in which the wellbeing of human beings could be harmed *and* it is not yet certain that this harm will occur for the technology is still in the research and development phase.

However, looking at the riskiness of technologies in this respect also seems limited in two ways that have to be kept in mind in the remaining of this article. Firstly, it needs to be remarked that a term like “risk” is mostly applicable to individuals. Technologies, however, may have effects on groups of people, as Janet Malek and Loretta Kopelman show in relation to DNA diagnostics which concern genes which people share. DNA tests can show that a specific group is more susceptible to a specific disease, such as Native Americans are to alcoholism. This genetic knowledge can lead to stigmatization and discrimination of that group. But these effects are hard to qualify as “harms” since groups lack the body and mind that is capable of being harmed. (Malek and Kopelman 2007) Words such as “risk” and “harm” are thus more apt to distinguish impacts on individuals, than they are to talk about effects on groups of people.

This is also a relevant point for the technology discussed in this article. While part of the possible future impacts of this technology will be understandable at an individual level, such as influences on capabilities such as life, health, imagination emotion, practical reason etc, others can only come into view if people are perceived

⁵See Oosterlaken, Ilse (forthcoming). “Design for Development; A Capability Approach”. In: *Design Issues* (accepted for publication on November 11th, 2008).

as *members of a group*. These effects are hard to understand in a risk-vocabulary; to do so demands an imaginative extension of the meaning of terms like “risk” and “harm”.

Secondly, the term “risk” may simplify the view of *how* new technologies influence people’s lives. Terms like “risk” and “harm” suggest that the relation between the introduction of a technology and its impact on human life is linear. But sometimes the relation between the introduction of a technology and its effects on people is not so straightforward. The stigmatization and discrimination that DNA-tests produce for Native Americans, for example, depend not only on the technology; rather, these effects are co-shaped by relations between population-groups prior to the introduction of DNA-tests in the US. DNA-tests are not the sole cause of stigmatization and discrimination of Native Americans; the availability of these tests *intensified and altered* a problem that was already there.

This will also be the case in the case-study that is presented here. Imagining the “impacts” that an acousto-optic monitoring device for blood may have on the quality of life of human beings also implies knowledge of the contextual characteristics in which this technology will be introduced. Without such contextual knowledge it becomes hard to imagine in a reliable way how people’s lives will be changed, and how they are likely to evaluate that change. Contextual knowledge is therefore a prerequisite for the formation of an “educated imagination” about a technology’s impact on the quality of human life.

These limitations have to be kept in mind in the discussion of this case-study, which will (1) imaginatively anticipate the changes in the capabilities that an acousto-optic monitoring device could produce, but will also (2) pay attention to possible group-effects, and (3) will try to come to grips with these changes by means of a study of the context for which the technology is intended.

2 The Acousto-Optic Monitoring Device for Blood

For an adequate anticipation of the interplay between a technology and a context, and the effects it will have on how human beings are able to conduct their lives, one needs to become acquainted with the technology first. Research into the acousto-optic non-invasive monitoring device for chemical substances in the blood, such as oxygen, glucose and cholesterol, is the most recent example of the search for a non-invasive monitoring method which began already in the nineteen thirties, when the first non-invasive instruments to measure oxygen in the blood were built. This research got accelerated during the Second World War; it was part of a project to investigate the oxygen level in the blood of pilots during fights at high altitudes, who frequently lost consciousness. After the war biophysicist E.H. Wood succeeded in constructing the first quantitative method to monitor oxygen levels in the blood, but this was only used in laboratories for it was not yet practical for use on patients. It took until 1972 when a simplification of Wood’s instrument was developed, and it was ready for use in a hospital context by 1983. This instrument was called the

“pulse oximeter”, and is nowadays adopted into standard anaesthesia practice in many countries. (Aoyagi 2003) The pulse oximeter is the clip patients get on their finger when they undergo surgery, and which monitors the oxygen-level in their blood. Next to surgery it is also widely used in recovery, emergency units and in intensive care units. Outside the hospital it is used in aviation, or by mountain-climbers, who need to monitor the oxygen level in their blood at high altitudes.

The pulse oximeter is an optical technique: it uses a light beam of infrared laser light, which points at part of the human body, most often a finger or an earlobe. It aims especially at arterial blood. When that blood is rich with oxygen it has a light colour, but when it contains little oxygen it is dark red; accordingly, blood rich with oxygen absorbs a lower amount of light, than the dark blood that contains little oxygen which absorbs a lot of light. The absorption degree of the light indicates the amount of oxygen in the blood, therefore the pulse oximeter is able to notice a critical oxygen-level before clinical signs are apparent. Since the pulse oximeter focuses on the oxygenation of *arterial* blood which has a pulse, the instrument is called “pulse oximeter”.

The pulse oximeter, however, is imprecise, because the light beam does not only go through the vessel –which is responsible for the measurement – but also through tissue, which strongly scatters the light. The pulse oximeter is unable to correct for this imprecision. This is one of the reasons why scientific engineers at the University of Twente engaged in research into a new technique, which uses sound as well as light, to overcome this problem. Research into this technology is still in a very early stage of development, meaning that it has only been researched *in-vitro* on a simulation of tissue, interestingly called a “phantom”. In the test-set-up the phantom is made by repetitive freezing and defrosting of an intralipid solution, which by that procedure acquires the substance of a white pudding that has light scattering characteristics that are similar to human tissue. In this white pudding a tube is inserted with coloured ink, which is the stand-in for the blood vessel. On one side a light beam with a well-defined colour is pointed at the phantom with the ink-tube, but on another side an acoustic transducer is pointed precisely at the tube. The ultrasound manipulates the movement of the photons (the light-particles) that transgress the tube, and allows to distinguish from the totality of light that leaves the body again the photons that go through the vessel from the photons that are scattered by tissue, thus allowing to focus on the absorption-level of only those photons. (Bratchenia et al. 2008)

This acousto-optic technology carries with it the promise to improve the results of measurements with the pulse oximeter, but its uses may also be extended in the future so that it is able to measure non-invasively other chemical substances in the blood, such as glucose or cholesterol, using other appropriate light colours. With the future exploration of these broader possibilities for the technology, this research builds forth on a wide field of research into techniques which aim to monitor in a semi-invasive or non-invasive way the glucose level in the blood of diabetes patients. Among these techniques are optical techniques such as optical sensors, but also other acousto-optical techniques. (Sieg et al. 2005; Larin et al. 2002; Zhao and Myllylae 2002)

The acousto-optic monitoring device may or may not become a usable technology. That depends on how successful the research will be. It is possible that this technology will be very successful and will deliver its promises, but experience teaches that many technologies will not succeed to take the step from the controlled environment of the laboratory to the much more complex reality of human bodies. However, also when the acousto-optic monitoring device does not become useable, the technology may become operational in other ways. In the near future, for example, research is planned into a connection with another technology that is being researched at the University of Twente, called photoacoustic mammography and which is intended for the non-invasive detection of breast cancer. (Manohar et al. 2007) Photoacoustic mammography also uses sound as well as light, but it is an imaging technique: it images excessive growth of blood vessels around a lump in the breast, which is an indication that it is a tumour.⁶ A connection between acousto-optics and photoacoustic mammography could make it possible to offer a more precise diagnosis to cancer patients. While photoacoustics is able to image extra vessel-growth around tumours, acousto-optics could determine the oxygenation of that blood. Blood with little oxygen indicates that the tumour grows fast, for it withdraws a lot of oxygen from the blood, while blood rich with oxygen is indicative of a slow growing tumour. Information about speed of growth could therefore make it possible to offer more precise diagnostic information than is available at present.

It seems worthwhile to anticipate the possible positive and negative ways in which an acousto-optic monitoring device could influence human life, for it may develop into a useable technique. But if it doesn't, it may become integrated into another technology such as photoacoustic mammography, which is already in a much more developed stage of development and will be tested on patients this year (2009). Furthermore, apart from the fact that it may be compatible with other techniques, there are also similarities between them. Its congeniality with other emerging sensing techniques and acousto-optic and photoacoustic technologies –which are researched all over the world⁷ – is an indication that these other technologies may share some of the problems that the acousto-optic monitoring device has, and which I will discuss in the following section.

3 Acousto-Optics and Dark Skin

There are many ways in which the acousto-optic instrument could affect the quality of human life. Here I will focus mostly on one example; namely, the different ways it may function on people with different skin colours. I first thought about

⁶ This process of vessel-formation around tumours is termed “angiogenesis”. (Carmeliet and Jain 2000). I provide for a more extensive case-description of photoacoustic mammography in van der Burg 2009.

⁷ It is important to realize that the field of optical technologies is broad. See, next to earlier mentioned articles about photoacoustics: Tromberg et al. 2000; Pogue et al. 2001. Articles about acousto-optic imaging techniques: Wang 2003; Lev and Sfez 2003.

the possibility that this technology could function less well on people with dark skin – meaning dark African or Indian skin, not the lighter skin-tones seen in Arabic or Asian countries – in the first phase of my embedded ethical research when the scientific engineers introduced me to their research-topic and their laboratory set-up. They explained to me that the measurement of the technology depended on the absorption of light by colour; also, they explained to me that in order to find out whether this technology worked they simplified reality in the laboratory, thus keeping the substitute of human tissue – the phantom – white while allowing only the content of the vein to be coloured. In relation to these explanations, different skin colours seemed to pose an obvious problem at least in the initial phases of the research. But the scientific engineers convinced me that they could eliminate this problem in a later research-phase, for it had also been solved in recent versions of the pulse oximeter, the most successful ancestor of their technology. However, during conversations with anesthesiologists in hospitals I found out that the pulse oximeter does not always work well on dark skin. These remarks drove me to search for literature about the pulse oximeter regarding this problem.

Interestingly, articles that focus on the technology of pulse oximetry, which are published in scientific journals about optical techniques, rarely mention skin colour at all in relation to the technology. For example Takuo Aoyagi, who collaborated in the development of the useable pulse oximeter that was realized in 1972, and began to be widely used in hospitals in 1983, does not mention skin-pigmentation as a problem when he discusses the history of the device, nor does he mention it in his inventory of the problems for its future. (Aoyagi 2003) And of the many articles that report about tests of the pulse oximeter on patients, only few take dark skin into consideration. Some of them report no alarming results. For example, two large-scale studies on 380 dark and white skinned subjects reported no significant pigment-related errors at a normal oxygen-level; “normal” meaning that haemoglobin in the blood – which is the oxygen carrier – contains between 95 and 99% oxygen. (Adler et al. 1998; Bothma et al. 1996) Many smaller-scale studies, however, reported errors in pulse oximeter readings in cases when the skin is coloured. Some of these studies were not especially aiming to study the performance of the pulse oximeter on dark skin, but reported that nailpolish, ink, henna or meconium (in neonates) is able to interfere with the pulse oximeter readings (Coté et al. 1988; Battito 1989; Goucke 1989; Johnson et al. 1990). There were also other studies which compared the performance of different kinds of pulse oximeters, and reported casually – as if it were just a “side-effect” – that they also produced different measurements on dark-skinned patients compared to those on light-skinned ones. (Cahan et al. 1990 and Seweringhaus and Kelleher 1992).

There are also studies that concentrate especially on skin-colour differences, and noticed a variation in pulse oximeter readings that is significant to medical decision-making. Jubran and Tobin (1990), for example, tested 54 critically ill patients who were ventilator-dependent and found out that the measurements of the pulse oximeter needed to be read differently in dark-skinned patients than in light-skinned ones. In patients with a light skin tone a pulse oximeter measurement that indicated that the blood contained 92% of oxygen could be considered safe, but dark skinned

patients with the same measurement suffered serious hypoxemia (lack of oxygen). In dark skinned patients the pulse oximeter needed to show an oxygen saturation of 95% to be considered “reliable”. Another test on 33 patients belonging to different ethnic groups – Indian (5), Malay (6) and Chinese (22) – by Lee et al. (1993) also indicates that the amount of oxygen is overestimated in people with dark skin, especially at low oxygen saturations in the blood. Misreadings in the dark-skinned Indian patient-group increased most when the oxygenation in the blood dropped.⁸ Ries et al. (1989) reported similar findings from a study on 187 patients for the pulse oximeter used in the ear: this type of oximeter produced a lot of technical problems often resulting in no reading at all on the darkest skinned test-subjects, and if it did work on these patients, it produced less accurate measurements.

These studies are all somewhat dated. A newer generation of pulse oximeters may have overcome these difficulties. This is what Gerard Coté claims (Coté 2001, p. 3). He argues that while earlier monitors were frustrated by a large list of problems, among which skin pigmentation, this is no longer a problem for a new generation of oximeters: now pulse oximeters use two wavelengths, which allow to take the ratio of the pulse and the total transmitted red light, and divide it by the same ratio for infrared light. The result should be dependent only on arterial oxygen-saturation, which should make pulse oximetry independent of skin colour.

While this technological explanation seems sound to scientific engineers –it is convincing to the researchers at the University of Twente – two small more recent patient-studies which were carried out on six different new brands of pulse oximeters, report that they are still not working accurately, especially not on dark skinned patients whose oxygen level in the blood drops below the normal.⁹ (Bickler et al. 2005; Feiner et al. 2007) Bickler et al (2005) report the results of a test on 21 test-subjects, 11 of whom had a dark skin colour (African American) and 10 a light Caucasian skin tone. During the experiment a measurement was taken with a “normal” oxygen level, and then the oxygen in the blood was lowered by means of breathing air-nitrogen-carbon dioxide mixtures through a mouthpiece. When the saturation of oxygen in the blood lowered, the bias in the pulse oximeter’s readings on dark skinned patients increased. In people with oxygen-levels beneath 80% a bias up to 8% was perceived, which can be significant in situations during surgery, in the emergency room, or for example in people with heart diseases who have a stable lower level of oxygen in the blood. In the study by Feiner et al. (2007) 36 test-subjects of different skin tones (ranging from white to Hispanic, Asian and African-American) were studied, among which were 19 males and 17 females. Next to skin colour this study also focused on gender differences in measurement, which affects finger-geometry. Here a deviation up to 5% was measured up until a saturation of 75% in dark skinned patients.

⁸The types of pulse oximeters checked here were Nellcore, Simed and Critikon.

⁹The pulse oximeters tested were Nonin, Masimo and Nellcor instruments in the last study (Feiner et al. 2007) The earlier study tested the Necor N-595 clip-on sensors and for Nonin Onyx and Novametrix 513. (Bickler et al. 2005).

In sum, while the skin-colour problem was thought to be technically solved, patient-tests give reason to think that this conclusion at least needs to be nuanced. Most pulse oximeters seem to be calibrated using light-skinned individuals, with the assumption that skin colour does not matter. But the studies mentioned above show that pulse oximeter readings need to be corrected by in vivo comparisons of oximeter readings in patients with different skin pigments. This is reason for several researchers who carried out such patient-studies to call for caution: some suggest that warnings about their bias on dark skinned patients should be printed on the instrument (Jubran and Tobin 1990, p. 1420), others argue that correction tables should become standard in hospitals or even that technical adjustments should be made which make it possible to adjust the instrument to the skin-tone of the patient. (Bickler et al. 2005, p. 717)

While all these suggestions are helpful solutions to the problem and could –if they would be communicated to technicians or anaesthetists, and would be taken seriously by them¹⁰ – eventually solve the problem for pulse oximetry, it is striking that 37 years since its invention the pulse oximeter is still not able to function properly on dark skin *and* that the problem acquires so little attention in technical articles. The scientific engineers from the University of Twente explained to me that the reason for that could be that the pulse oximeter was created by mostly light-skinned Caucasian and Asian people, and is funded by institutions in countries where light-skinned people dominate and will thus be inattentive to skin colour problems. Next to that, they argue that skin colour cannot be paid attention to by scientific engineers because of the rigid phase-structure that technological research needs to respect if it is to be successful: this means that researchers first deal with simplified versions of reality in the laboratory and will only consider complications such as “skin colour” when their technology is tested on patients and turns out to have troubles on dark-skinned patients. Abandoning this phase-structure of research would mean, according to the scientists, that it is impossible to do research that is conducive to the development of an actual technology.

While these are both relevant and understandable explanations for the reluctance of scientific engineers to pay attention to literature about test-results in an early phase of their research, it also shows why a new technology –such as acousto optics – risks to have the same flaws as the pulse oximeter. If scientific engineers do not know about the pulse oximeter’s poor performance on people with dark skin at low oxygen levels, they lack important knowledge that may be informative to their research-questions and test set-ups, including the set-up of the test phase on patients. Such patient-tests do not show by themselves that a technology performs

¹⁰This is not a matter of course. Communication between researchers and hospitals remains limited, as well as the communication between researchers and producers. Next to that, it is striking that a lot of information never leads to action: it is for example known that it is harder to withdraw blood from dark skinned people, because the vein cannot be found easily. This makes withdrawing blood for dark people a lot more painful and distressing (especially for children). While there are technological possibilities to solve this problem, there has to be a researcher who takes an interest in it, to be able to solve it.

poorly on dark skinned patients; tests can only offer that information if – prior to the test – information is gathered on the skin tones of the test-subjects, and performance on different skin colours is included among the research questions. Setting the patient-tests up in this way means that scientists need to hypothesize *before-hand* that there might be a skin-colour related problem. They are only able to do that if they have information, such as the information provided here about the pulse oximeter's problems with dark skin.

Of course, acousto-optics does not function in exactly the same way as the pulse oximeter, for next to light it also uses sound. Ultrasound is used to alter the amplitude with which the light goes through the vessel, and allows to distinguish the photons that traverse the vessel from the other light particles. This means that the measurement depends on the photons that go through the vessel. The rest of the light is ignored in the measurement. However, if part of the light is absorbed at the level of the skin, there are less photons left that go through the vessel. Furthermore, the measurement depends on the light-signal that comes back out of the body. So, the light has to go through the skin twice – in and out – meaning that the skin colour will absorb part of that light twice too, meaning that it is not sure whether a measurement will be possible. This seems especially so when the oxygen-level in the blood is low, for in that case the little light that succeeds to go through the skin and reaches the vessel may be absorbed by the darkly coloured blood, which leaves almost no light to go back out of the body, through the skin, and allow a measurement.

Whether it is more difficult, or impossible, to make a measurement on dark skinned people with an acousto-optic device, of course needs to be researched. But my discussion at least shows that there is a potential problem here that needs attention. The problem of skin-colour is especially relevant because acousto-optics also aims to bring about a non-invasive monitoring technique for diabetes patients. While this technology will probably use different wavelengths of light to study the glucose-level, which may alter its performance on dark skin, it is important to note that skin-colour is a relevant research-factor in the area of glucose-monitoring for diabetes. Information about contexts for which this technology is intended shows why that is so. In the Netherlands, for example, diabetes patients often have a dark skin. According to the National Compass for Public health, which gives information about the occurrence of diseases in the Dutch population, diabetes mellitus occurs more frequently among populations from Suriname, Morocco and Turkey, than among people of Dutch decent. Especially Hindu people from Suriname –who have a very dark skin – have a relative high chance of developing diabetes: 37% of the population older than 60 has the disease. The presence of diabetes among people from Turkey and Morocco also lies 3–6 times higher than among people from Dutch decent.¹¹ That means that at least within the Netherlands, diabetes patients will to a

¹¹ Reasons for this difference are hard to give, but it is thought to be explained by deprivation during youth, and the more frequent occurrence of obesity among these populations. English sources on which these findings are based are for example: Middelkoop et al. 1999; Weijers et al. 1998. For Dutch readers: see the site of RIVM http://www.rivm.nl/vtv/object_document/01261n17502.html

large extent be people with a darker skin.¹² But studies outside the Netherlands have also pointed out that the amount of diabetes patients is high in certain areas where people with very dark skin colours live, the most well-known example being India. (Ranachandran et al. 2001) This empirical information shows why it is important for researchers and developers of non-invasive monitoring techniques for glucose to pay attention to skin colour.

4 Possible “Risks” of the Acousto-Optic Monitoring Device

The role I am proposing for embedded ethicists is to imagine the “risks” that new technologies might bring about in the future for people’s quality of life, and bring that information to the laboratory so that it can inform research decisions. This imagination, however, needs to be an “educated imagination”: it is formed on the basis of information about the new technology, as well as other related technologies. It also needs information about the contexts for which the technology is eventually intended. This information allows to imagine in a reliable way what types of hard and soft impacts this technology is likely to have on human (social) life. Here I will give an example of such an imaginary endeavour.

If acousto-optics succeeds to make possible a more precise and adequate non-invasive measurement of oxygen than the pulse-oximeter is able to deliver, it is clear why it would be desirable to have such an instrument. It will improve the means to monitor the life and health of patients during surgery. If the acousto-optic monitoring device turns out to be able to measure glucose too, it would even have more attractive impacts. In that case, it would offer a less painful and less inconvenient way to check glucose than the invasive check-ups that are currently the standard for diabetes patients. Also, it enables diabetes patients to check their glucose level frequently, which allows them to keep it more balanced and that may lead to the development of fewer complications. If research into the acousto-optic monitoring device is successful, it could therefore contribute to “hard impacts” such as the diminishment of pain during the glucose-checks, as well as a decrease of the amount and seriousness of the complications that are frequent symptoms in people with diabetes, such as problems with eyesight, kidneys, peripheral nerves, heart and blood vessels. This of course fosters the capabilities of life, bodily integrity and health.

Furthermore, the acousto-optic device could also deliver soft impacts, which depend on specific characteristics of the context for which it is intended. During a focus group that I organized twelve Dutch diabetes patients pointed out how their life is affected by the frequent invasive glucose-checks at this moment, which helps to imagine what changes will occur if the check-ups become non-invasive. These diabetes patients check themselves between 4 and 10 times a day, which is a painful and troublesome procedure. A student describes, for example, how she had to overcome a fear of needles to be able to carry out her daily glucose-checks. While

¹²People from Morocco and Turkey usually have lighter skin colours than Hindu people from Suriname who generally have dark brown skin.

she handles her fear now, she describes always feeling repulsed when she has to insert the needle. This fear is not shared by the other interviewed patients, but they all report that they often feel unwilling to interrupt their activities to check their glucose, and express regret at missing things: such as courses, part of a lively conversation during lunch with colleagues, or even part of work. A man, who works as a sculptor, says he fails to check his glucose when he is working, “because it is too much of a hassle to get clean and perform the check-up.” Another man who works in construction agrees with him and explains that the time-consuming procedure to clean his fingers adequately and do the glucose-check has led to tensions between him and his colleagues. “I know it is good for my health”, he explains. “But not all of my colleagues understand: they feel I abandon them and they have to do the hard work.”

These experiences indicate that a non-invasive acousto-optic monitoring device could bring about “soft impacts” such as liberation from the fear of needles and the inconvenient interruption of daily activities, but it would also enable people to engage more thoroughly in their work or studies and enjoy and sustain more adequately relationships with colleagues and friends. The technology would thus help to develop capabilities of affiliation with others, and of emotions, and likely also of practical reason. Practical reason is the capacity to form a conception of the good and to engage in critical planning of one’s life: since the acousto-optic monitoring device liberates people from the burden of frequent time-consuming invasive tests, it could enable people to plan their life and actions more freely. Having to take frequent invasive tests could mean that a patient will not choose a career that involves getting dirty hands; and it puts constraints on what actions an individual is able to do in a day. If the test is non-invasive, diabetes patients could have more freedom to deliberately plan their lives, instead of having to organize it around their disease.

This catalogue of hard and soft impacts that an acousto-optic monitoring device could deliver is attractive. However, the skin-colour story points out that these impacts may be realized only for light skinned people. The above mentioned capabilities as life, bodily integrity, health, emotion, affiliation and practical reason may therefore be fostered in white skinned patients, and not in people with dark skin. This would mean, of course, that dark skinned patients are not enabled to realize their “good life” as well as light-skinned people are.

Thus, the technology is likely to produce an inequality between the lives of light skinned diabetes patients and dark skinned ones, which was not there before. This could make dark skinned patients vulnerable in new ways. It would mean, for example, that dark skinned patients – unlike white ones – have to continue to interrupt their daily activities for the time-consuming tests, while white skinned people with diabetes manage to check glucose quickly. This offers more freedom to whites to form their lives and relationships in the way they desire than blacks. Tensions that can come about between people, such as colleagues, because diabetes patients have to regularly retreat from their activities to perform their invasive tests, will then be reserved exclusively to people of dark skin colours. And it is quite possible that the outside world of people without diabetes will have more difficulty understanding the time and effort that dark skinned patients have to invest to keep their glucose at an even level, if their white skinned fellow-patients fulfil the same task much quicker.

This likely also affects the emotions. In a more extensive explanation of the catalogue of capabilities that Nussbaum gives in a later work, she pays special attention to the emotions. Here she identifies emotion as a capability “(..) to have attachments to things and people outside ourselves; to love those who love and care for us, to grieve at their absence; in general, to love, to grieve, to experience longing, gratitude and justified anger.”(Nussbaum 2000, p. 79) The capability to have emotions here refers to a capacity to relate to other people, and to express the emotions that one experiences. If we imagine the impacts of an acousto-optic instrument which works on light skinned people, but not on people with dark skin pigment, we see that it affects emotions in precisely this way: it makes relations between people more difficult. If dark skinned patients, unlike light-skinned ones, remain tied to invasive tests, they have less control over their life-plan and their day-planning. And this may elicit irritation of other people, who fail to understand why they have to retreat so frequently.

In addition, relations between diabetes patients may alter, because they no longer share the same experiences with the disease. The difficulties dark-skinned patients experience because they have to take the invasive tests are no longer the same experiences as white skinned patients have. This drives the two groups apart. It will demand an extra effort for light-skinned patients to be able to notice that their dark-skinned fellow diabetes patients are vulnerable in other ways than they are. And of course it is very common that people do not take the effort to imagine themselves in someone else’s shoes.

Altered emotions are also likely to impact on people’s value judgments. In her book about emotions *Upheavals of thought* Nussbaum defends the view that “Emotions (..) involve judgments about important things, judgments in which, appraising an external object as salient for our own wellbeing, we acknowledge our own neediness and incompleteness before parts of the world that we do not fully control.” (Nussbaum 2001, p. 19) Nussbaum here states that emotions are forms of evaluative judgment that ascribe to things and persons outside of a person’s control great importance for that person’s own flourishing.

In the imaginative exploration of the future that is offered here we have seen that a new technology such as the acousto-optic monitoring device is able to alter vulnerabilities: it takes some vulnerabilities away. And it probably takes them away for some people, but not for others. This means that different people are likely to experience different emotions, since they are in differing ways “incomplete”. The needs of blacks will not be the same as the needs of whites, and consequently blacks and whites will judge differing objects to be salient to their wellbeing. This difference has of course effects on how people choose to lead their own lives, but it also is able to affect society. People who have difficulty imagining the lives of others, will often fail to consider the difficulties of others when they deliberate about their own decisions. It is in such a way that for example the pulse oximeter got calibrated on light-skinned people; the engineers who did that simply did not sufficiently imagine difficulties that could arise on dark skin. Because of the decisions of many individuals, technologies are produced that affect the society in which individuals live their lives.

5 Concluding Remarks

In this imaginary anticipation of the future of the acousto-optic monitoring device for oxygen and glucose in the blood, I have tried to track its possible effects on human capabilities. I have attempted to show that there's a reasonable risk that this technology does not work, or that it works less well, on dark skinned people. This difficulty is likely to produce many effects on people, the most important one being that it will make it more difficult for dark and white skinned patients to relate to each other's vulnerabilities. This is bound to affect their judgments about what is valuable and worth pursuing in their personal, as well as professional or public life.

This "educated imagination" about the possible ways in which an acousto-optic monitoring device could affect people can be used as input in conversations with scientific engineers. The purpose of that would be to broaden the scope of scientists' deliberations about their research. Usually the imaginations of scientific engineers about the future of their technology depend largely on their technological knowledge, which gives them a broad and complicated field of research-questions to study. However, the more "ethical imagination" explored in this article could raise some additional research questions, or alter research priorities, or it could offer alternative views on how part of the research – such as the testing phase – should be conducted. While it is understandable that scientific engineers have to limit the scope of their attention to be able to acquire the level of specialization that is needed to do their research, they do contribute to technologies that potentially change people's lives. It therefore seems worthwhile that others – such as ethicists – do the extensive work that is needed for the development of a broader view of the future, which includes also the effects on a rich variety of users. Such an "ethical imagination" offers insights that are hopefully able to draw the scientists attention to some aspects of their technology that are worth their attention, and that they previously did not consider.

Acknowledgments With thanks to Sabine Roeser for the invitation to contribute to this book, as well as for her comments on the first draft of this article. I am also thankful to the Biophysical Engineering Group at the University of Twente for their generous and open cooperation with the embedded ethical research that lead to this article, and to the volunteers from the Diabetes Association Netherlands (Diabetesvereniging Nederland) who participated in the focus group. I also want to thank Rob Kooyman for his comments on an earlier version of this article. Finally I would like to thank NWO for funding the embedded ethical research that resulted in this article.

References

- Adler, J. N., L. A., Hughes, R., Vivilecchia, and C. A., Camargo, Jr. 1998. Effect of skin pigmentation on pulse oximetry accuracy in the emergency department. *Academic Emergency Medicine* 5: 956–970.
- Aoyagi, T. 2003. Pulse oximetry: Its invention, theory and future. *Journal of Anesthetics* 17: 259–266.
- Battito, M. F. 1989. The effect of fingerprinting ink on pulse oximetry. *Anesthesia & Analgesia* 69: 265.

- Bickler, P. E., J. R., Feiner, and J. W., Severinghaus. 2005. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 102: 715–719.
- Bothma, P. A., G. M., Joynt, J., Lipman, H., Hon, B., Mathala, J., Scribante, and J., Kromberg. 1996. Accuracy of pulse oximetry in pigmented patients. *South African Medical Journal* 86: 594–596.
- Bratchenia, A., R., Molenaar, and R. P. H., Kooyman. 2008. Feasibility of quantitative determination of local optical absorbances in tissue-mimicking phantoms using acousto-optical sensing. *Applied Physics Letters* 92: 113901.
- Cahan, C., M. J., Decker, P. L., Hoekje, and K. P., Strohl. 1990. Agreement between non-invasive oximetric values for oxygen saturation. *Chest* 97: 814–819.
- Carmeliet, P., and R. K., Jain. 2000. Angiogenesis in cancer and other diseases. *Nature* 407: 247–257.
- Collingridge, D.. 1980. *The Social Control of Technology*. London: Pinter Publishers.
- Coté, C. J., E. A., Goldstein, W. H., Fuchsmann, and D. C., Hoaglin. 1988. The effect of nail polish on pulse oximetry. *Anesthesia & Analgesia* 67: 683–686.
- Coté, G. L.. 2001. Noninvasive and minimally-invasive optical monitoring technologies. *American Society for Nutritional Sciences* 131: 1596s.
- Feiner, J. A., J. W., Severinghaus, and P. E., Bickler. 2007. Dark skin decreases the accuracy of pulse oximeters at low oxygens: The effects of oximeter probe type and gender. *Anesthesia & Analgesia* 105: 18–23.
- Fisher, E.. 2007. Ethnographic invention: Probing the capacity of laboratory decisions. *Nanoethics* 1: 155–165.
- Goucke, R.. 1989. Hazards of henna. *Anesthesia & Analgesia* 69: 416–417.
- Hansson, S. O.. 1996. Decision making under great uncertainty. *Philosophy of the Social Sciences* 26: 369–386.
- Hansson, S. O.. 2003. Ethical criteria of risk acceptance. *Erkenntnis* 59: 291–309.
- Johnson, N., V. A., Johnson, J., Bannister, and H., McNamara. 1990. The effects on maconium on neonatal and fetal reflectance pulse oximetry. *Journal of Perinatal Medicine* 18: 351–355.
- Jubran, A., and M. J., Tobin. 1990. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. *Chest* 97: 1420–1425.
- Larin, K. V., M. S., Eledrisi, M., Motamedi, and R. O., Esenaliev. 2002. Non-invasive blood glucose monitoring with optical coherence tomography: A pilot study in human subjects. *Diabetes Care* 25: 2263–2267.
- Lee, K. H., K. P., Hui, W. C., Tan, and T. K., Lim. 1993. Factors influencing pulse oximetry as compared to functional arterial saturation in multi-ethnic Singapore. *Singapore Medical Journal* 34: 385–387.
- Lev, A., and B., Sfez. 2003. In vivo demonstration of the ultrasound-modulated light technique. *Optical Society of America* 20: 2347–2354.
- Lynch, W. T., and R., Kline. 2000. Engineering practice and engineering ethics. *Science, Technology & Human Values* 25: 195–225.
- Malek, J., and L. M., Kopelman. 2007. The well-being of subjects and other parties in genetic research and testing. *Journal of Medicine and Philosophy* 32: 311–319.
- Manohar, S., S. E., Vaartjes, J. C. G., van Hespden, J. M., Klaase, F. M., van den Engh, W., Steenbergen, and T. G., van Leeuwen. 2007. Initial results of in vivo non-invasive cancer imaging in the human breast using near infrared photoacoustics. *Optics Express* 15: 19.
- Middelkoop, B. J. C., S. M., Kesarlal-Sadhoeran, G. N., Ramsaransing, and H. W., Struben. 1999. Diabetes Mellitus among South Asian inhabitants of the Hague: High prevalence and an age-specific socioeconomic gradient. *International Journal of Epidemiology* 28: 1119–1123.
- Nussbaum, M. C.. 2000. *Women and Human Development; the Capabilities Approach*. Cambridge/New York: Cambridge University Press.
- Nussbaum, M. C.. 2001. *Upheavals of Thought; the Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Nussbaum, M. C. and A. Sen (ed.). 1993. *The Quality of Life*. Oxford: Clarendon Press.

- Pogue, B. W., S. P., Poplack, T. O., McBride, W. A., Wells, K. S., Osterman, U. L., Osterberg, and K. D., Paulsen. 2001. Quantitative hemoglobin tomography with diffuse near-infrared spectroscopy: Pilot results in the breast. *Radiology* 218: 261–266.
- Ranachandran, A., C., Snehalatha, A., Kapur, V., Vijay, V., Mohan, A. K., Das, P. V., Rao, C. S., Yajnik, K. M., Prasanna Kumar, and I. D., Nair. 2001. High prevalence of diabetes and impaired glucose tolerance in India: National urban diabetes survey. *Diabetologia* 44: 1094–1101.
- Ries, A. L., L. M., Prewitt, and J. J., Johnson. 1989. Skin color and ear oximetry. *Chest* 96: 287–290.
- Rip, A., J. Schot, and T. J. Misa (eds.). 1995. *Managing Technology in Society: The Approach of Constructive Technology Assessment*. London: Pinter.
- Selb, J., S., Lévêque-Fort, L., Pottier, and C., Boccara. 2001. 3D acousto-optic modulated-speckle imaging in biological tissues. *Physique appliquée/Applied Physics* 2: 1213–1225.
- Sieg, A., R. H., Guy, and M. B., Delgado-Charro. 2005. Non-invasive and minimally invasive methods for transdermal glucose monitoring. *Diabetes Technology & Therapeutics* 7: 174–197.
- Sweringhaus, J. W., and J. F., Kelleher. 1992. Recent developments in pulse oximetry. *Anaesthesiology* 76: 1018–1038.
- Teuber, A.. 1990. Justifying risk. *Daedalus* 119: 1–20.
- Tromberg, B. J., N., Shah, R., Lanning, A., Cerussi, J., Esponzoza, T., Pham, L., Svaasand, and J., Butler. 2000. Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy. *Neoplasia* 2: 26–40.
- van der Burg, S.. 2009. Imagining the future of photoacoustic mammography. *Science and Engineering Ethics* 15: 97–111.
- Weijers, R. N. M., D. J., Bekedam, and H., Oosting. 1998. The prevalence of type 2 diabetes and gestational diabetes mellitus in an inner city multi-ethnic population. *European Journal of Epidemiology* 14: 693–699.
- Xu, M., and L. V., Wang. 2006. Photoacoustic imaging in biomedicine. *Review of Scientific Instruments* 77: 261–266.
- Wang, L. V.. 2003. Ultrasound-mediated biophotonic imaging: A review of acousto-optical tomography and photo-acoustic tomography. *Disease Markers* 19: 123–138.
- Zhao, Z., and R. A., Myllylae. 2002. Photoacoustic blood glucose and skin measurement based on optical scattering effect. *Proceedings – Society of Photo-Optical Instrumentation Engineers* 707: 153–157.
- Zwart, S. D., I., van de Poel, H., van Mil, and M., Brumsen. 2006. A network approach for distinguishing ethical issues in research and development. *Science and Engineering Ethics* 12: 663–684.

Part III
Emotions as a Guide to Acceptable Risk

Emotion in Risk Regulation: Competing Theories

Dan M. Kahan

1 Introduction

Are emotions subversive of reason or essential constituents of it? Do they defeat realization of our ends by enfeebling our calculative faculties, inducing us to form deluded beliefs, and undermining our wills? Or do they perfect our rationality by supplying us with a capacity to perceive which states of affairs express our values, the motivation to pursue those conditions, and the power to imagine contingencies that threaten or advance them? These questions have long divided both philosophers and psychologists (Elster 1999). Competing answers contend with one another in law as well (Kahan and Nussbaum 1996).

Recent advances in the study of risk perception seem to furnish decisive evidence of emotion's antagonism to reason. A growing body of empirical research supplies compelling evidence of the critical role that emotions play in the apprehension of personal and societal dangers (Slovic et al. 2005). This role, according to the predominant understanding, is a heuristic one. Lacking access to sound empirical information, or the time and cognitive capacity to make sense of it, ordinary people conform their perceptions of risk to the visceral reactions that putatively dangerous activities evoke (Loewenstein et al. 2001). These snap judgments might serve individuals better than nothing, the conventional account suggests. But they don't serve individuals nearly as well as the type of considered, reflective assessment for which they are a substitute. A substantial body of writing in the field of risk perception documents the numerous ways in which affect-driven risk appraisals lead ordinary people, and their popularly accountable representatives, to take positions inimical to society's well-being. The remedy, according to this work, is to shield law from the distorting influence of emotion, primarily by delegating regulatory power to politically insulated experts, who can evaluate the costs and benefits of asserted hazards (nuclear power, genetically modified foods, handguns, etc.) in a deliberate and reasoned fashion (Sunstein 2005; Breyer 1993).

D.M. Kahan (✉)

Elizabeth K. Dollard Professor of Law, Yale Law School, New Haven, CT, USA
e-mail: Dan.Kahan@yale.edu

My goal in this essay is to challenge this position. I don't mean to raise any question about the demonstrated centrality of emotions to risk perception, but only about the prevailing interpretation of it. The conclusion that emotional appraisals are irrational is integral, I'll argue, to a model of risk perception that sees the positions people take toward putatively dangerous activities as reflecting their implicit (and usually skewed) weighing of instrumental costs and benefits. I will lay emphasis instead on an account that sees risk perceptions as embodying individuals' cultural evaluations of the meanings expressed by society's decision to tolerate or abate particular risks (Kahan et al. 2006). This model of risk perception, I'll argue, suggests that emotion functions not as a heuristic substitute for considered appraisals of information but rather as a perceptive faculty uniquely suited to discerning what stance toward risk best coheres with a person's values. Without the power this affective capacity supplies, it would be impossible for individuals to form rational cultural evaluations of risk. This account suggests that it would be a mistake, too, to seal off risk regulation from the influence of affect-driven risk appraisals or to assume that affect-driven appraisals cannot themselves be influenced by education and deliberation.

I will develop this argument in three steps. I will begin, in Section 2 of this essay, by describing three theories of risk perception, two of which treat emotion as essential to the cognition of risk. In Section 3, I will canvass empirical findings that bear on these alternative understandings of how emotion contributes to risk perception. Finally, in Section 4, I will examine what is at stake as a normative and prescriptive matter in the contest between these two conceptions of emotion in risk regulation.

2 Three Theories of Risk Perception, Two Conceptions of Emotion

The profound impact of emotion on risk perception cannot be seriously disputed. Distinct emotional states – from fear to dread to anger to disgust (Slovic 2000) – and distinct emotional phenomena – from affective orientations to symbolic associations and imagery (Peters and Slovic 2007) – have been found to explain perceptions of the dangerousness of all manner of activities and things – from pesticides (Alhakami and Slovic 1994) to mobile phones (Siegrist et al. 2005), from red meat consumption (Berndsen and van der Pligt 2005) to cigarette smoking (Slovic et al. 2005).

More amenable to dispute, however, is exactly *why* emotions exert this influence. Obviously, emotions work in conjunction with more discrete mechanisms of cognition in some fashion. But which ones and how? To sharpen the assessment of the evidence that bears on these questions, I will now sketch out three alternative models of risk perception – the rational weigher, the irrational weigher, and the cultural evaluator theories – and their respective accounts of what (if anything) emotions contribute to the cognition of risk.

2.1 The Rational Weigher Theory: Emotion as Byproduct

Based on the premises of neoclassical economics, the *rational weigher theory* asserts that individuals, over time and in aggregate, process information about risky undertakings in a way that maximizes their expected utility. The decision whether to accept hazardous occupations in exchange for higher wages, (Viscusi 1983) to engage in unhealthy forms of recreation in exchange for hedonic pleasure, (Philipson and Posner 1993) to accept intrusive regulation to mitigate threats to national security (Posner 2006) or the environment (Posner 2004) – all turn on a utilitarian balancing of costs and benefits.

On this theory, emotions *don't* make any contribution to the cognition of risk. They enter into the process, if they do at all, only as reactive byproducts of individuals' processing of information: if a risk appears high relative to benefits, individuals will likely experience a negative emotion – perhaps fear, dread, or anger – whereas if the risk appears low they will likely experience a positive one—such as hope or relief (Loewenstein et al. 2001). This relationship is depicted in Fig. 1.

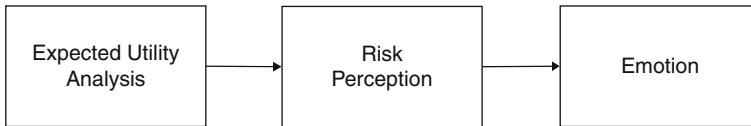


Fig. 1 The rational weigher theory of risk perception

2.2 The Irrational Weigher Theory: Emotions as Bias

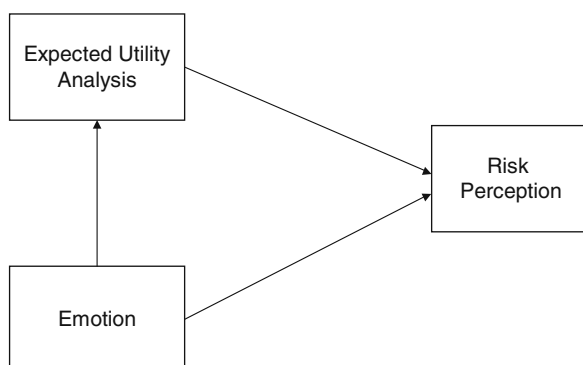
The irrational weigher theory asserts that individuals lack the capacity to process information that maximizes their expected utility. Because of constraints on information, time, and computational power, ordinary individuals must resort to heuristic substitutes for considered analysis; those heuristics, moreover, invariably cause individuals' evaluations of risks to err in substantial and recurring ways (Jolls et al. 1998). Much of contemporary social psychology and behavioral economics has been dedicated to cataloging the myriad distortions – from the “availability cascades” (Kuran and Sunstein 1998) to “probability neglect” (Sunstein 2002) to “overconfidence” bias (Fischhoff et al. 1977) to “status quo bias” (Kahneman 1991) – that systematically skew risk perceptions, particularly those of the lay public.

For the irrational weigher theory, the contribution that emotion makes to risk perception is, in the first instance, a heuristic one. Individuals rely on their visceral, affective reactions to compensate for the limits on their ability to engage in more considered assessments (Loewenstein et al. 2001; Slovic et al. 2004). More specifically, irrational weigher theorists have identified emotion or affect as a central component of “System 1 reasoning,” which is “fast, automatic, effortless, associative, and often emotionally charged,” as opposed to “System 2 reasoning,” which

is “slower, serial, effortful, and deliberately controlled” (Kahneman 2003, p. 1451), and typically involves “execution of learned rules” (Frederick 2005, p. 26). System 1 is clearly adaptive in the main – heuristic reasoning furnishes guidance when lack of time, information, and cognitive ability make more systematic forms of reasoning infeasible – but it remains obviously “error prone” in comparison to the “more deliberative [and] calculative” System 2 (Sunstein 2005, p. 68).

Indeed, according to the irrational weigher theory, emotion-pervaded forms of heuristic reasoning can readily transmute into bias. The point isn’t merely that emotion-pervaded reasoning is less accurate than cooler, calculative reasoning; rather it’s that habitual submission to its emotional logic ultimately displaces reflective thinking, inducing “behavioral responses that depart from what individuals view as the best course of action” – or at least would view as best if their judgment were not impaired (Loewenstein et al. 2001). Proponents of this view have thus linked emotion to nearly all the cognitive biases shown to distort risk perceptions (Fischhoff et al. 1977; Sunstein 2005). The relationship between emotion, rational calculation of expected utility, and risk perception that results is depicted in Fig. 2.

Fig. 2 Irrational weigher theory of risk perception



2.3 The Cultural Evaluator Theory: Emotion as Expressive Perception

Finally there’s the *cultural evaluator theory* of risk perception. This model rests on a view of rational agency that sees individuals as concerned not merely with maximizing their welfare in some narrow consequentialist sense but also with adopting stances toward states of affairs that appropriately *express* the values that define their identities (Anderson 1993). Often when an individual is assessing what position to take on a putatively dangerous activity, she is, on this account, not weighing (rationally or irrationally) her expected utility but rather evaluating the *social meaning* of that activity (Lessig 1995). Against the background of cultural norms (particularly

contested ones), would the law's designation of that activity as inimical to society's well-being affirm her values or denigrate them (Kahan et al. 2006)?

Like the irrational weigher theory, the cultural evaluator theory treats emotions as entering into the cognition of risk. But it offers a very different account of how – one firmly aligned with the position that sees emotions as constituents of reason.

Martha Nussbaum describes emotions as “judgments of value” (Nussbaum 2001). They orient a person who values some good, endowing her with the attitude that appropriately expresses her regard for that good in the face of a contingency that either threatens or advances it. On this account, for example, *grief* is the uniquely appropriate and accurate judgment for someone who values another who has died; *fear* is the appropriate and accurate judgment for someone who values her or another's well-being in the face of an impending threat to it; *anger* is the appropriate and accurate judgment for someone who values her own honor in response to an action that conveys insufficient respect. People who fail to experience these emotions under such circumstances – or who experience these or other emotions in circumstances that do not warrant them – lack a capacity of discernment essential to their flourishing as agents capable of holding values and pursuing them.

Rooted heavily in Aristotelian philosophy, Nussbaum's account is, as she herself points out, amply grounded in modern empirical work in psychology and neuroscience. Antonio Damasio's influential “somatic marker” account, for example, identifies emotions with a particular area in the brain (Damasio 1994). Persons who have suffered damage to that part of the brain display impaired capacity to recognize or imagine conditions that might affect goods they care about, and thus lack motivation to respond accordingly. They are perceived by others and often by themselves as mentally disabled in a distinctive way, as suffering from a profound kind of moral and social obtuseness that makes them incapable of engaging the world in a way that matches their own ends. If being rational consists, at least in part, of “see[ing] which values [we] hold” and knowing how to “deploy these values in [our] judgments,” then “those who are unaware of their emotions or of their emotional lacks” will necessarily be deficient in a capacity essential to being “a rational person” (Stocker and Hegeman 1996, p. 105).

The cultural evaluator theory views emotions as enabling individuals to perceive what stance toward risks coheres with their values. Cultural norms obviously play a role in shaping the emotional reactions people form toward activities such as nuclear power, handgun possession, homosexuality, and the like (Elster 1999). When people draw on their emotions to judge the risk that such an activity poses, they form an expressively rational attitude about what it would *mean* for their cultural world-views for society to credit the claim that that activity is dangerous and worthy of regulation, as depicted in Fig. 3. Persons who subscribe to an egalitarian ethic, for example, have been shown to be particularly sensitive to environmental and technological risks, the recognition of which coheres with condemnation of commercial activities that generate distinctions in wealth and status. Persons who hold individualist values, in contrast, tend to dismiss concerns about global warming, nuclear

waste disposal, food additives, and the like – an attitude that expresses their commitment to the autonomy of markets and other private orderings (Douglas 1966). Individualistic persons worry instead about the risk that gun control – a policy that denigrates individualist values – will render law-abiding citizens defenseless (Kahan et al. 2007a). Persons who subscribe to hierarchical values worry about the dangers of drug distribution, homosexuality, and other forms of behavior that defy traditional norms (Wildavsky and Dake 1990).

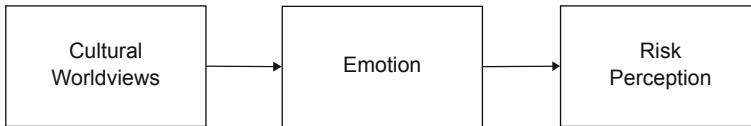


Fig. 3 The cultural evaluator theory of risk perception

This account of emotion doesn't see its function as a heuristic one. That is, emotions don't just enable a person to latch onto a position in the absence of time to acquire and reflect on information. Rather, as a distinctive faculty of cognition, emotions perform a unique role in enabling her to identify the stance that is expressively rational for someone with her commitments. Without the contribution that emotion makes to her powers of expressive perception, she would be lacking this vital incident of rational agency, no matter how much information, no matter how much time, and no matter how much computational acumen she possessed.

3 Empirical Evidence

3.1 *The Cognitive Priority of Emotion to Risk Perception*

Among the most important empirical studies on emotion and risk perception are those that demonstrate the cognitive priority of the former. Rather than conform their emotional appraisals of a putatively dangerous activity (say, nuclear power generation) to their assessment of its risks, individuals conform their assessments of its risks to their emotional appraisals (Alhakami and Slovic 1994).

This finding tells decisively against the rational weigher theory of risk perception. Because that theory assumes that individuals will rationally process information in a way that maximizes their expected utility, it doesn't supply any reason to believe that persons who have different emotional reactions toward an activity will form different factual beliefs about its risks and benefits (Loewenstein et al. 2001).

The cognitive priority of emotion to risk perception *is* consistent, however, with the irrational weigher theory. Under that theory, emotions directly influence risk perceptions direction as a heuristic, System 1 substitute for more reflective System 2 reasoning, and indirectly as a distorting force on individuals' processing of information.

The cultural evaluator theory also asserts that emotion exerts a cognitive influence on risk perception – not by distorting the processing of information, but by enabling individuals to perceive what stance toward risk *rationaly* expresses their cultural worldviews. Studies that tell us only that emotion is cognitively prior to risk perceptions, then are equally compatible with both the cultural evaluator theory’s conception of emotion as expressive perception and the irrational weigher theory’s conception of emotion as bias.

3.2 The Effects of Emotion on Information Processing

Another class of studies purports to identify particular characteristics of individuals’ risk perceptions that are plausibly viewed as evidence of the impact of emotion on information processing. Studies of this sort, however, also fail to resolve decisively the dispute between emotion as bias and emotion as expressive perception.

One feature of risk perception said to bear the signature of emotion is the unwillingness of individuals to adjust their decisions about the acceptability of risks to changes in information about their probability (Loewenstein et al. 2001). System 2 reasoning requires not only that people form unbiased assessments of the magnitude of risks and benefits, but also that they appropriately combine them to determine the expected utility of forgoing or forbearing them. That doesn’t happen when people are emotional. Instead they fail to discount a potential harm by its improbability – the phenomenon of “probability neglect” – because “when intense emotions are engaged, people tend to focus on the adverse outcome, not on its likelihood” (Sunstein 2005, p. 64). By the same token, when people “anticipate a loss of what [they] now have, [they] can become genuinely afraid, in a way that greatly exceeds [their] feelings of pleasurable anticipation when [they] look forward to some supplement to what [they] now have” (Sunstein 2005, p. 41). The result is “status quo” bias, the disposition to refrain from action that entails some risks but that nonetheless has a positive expected value (*ibid.*). Alternatively, positive emotions – such as hope or pride – can lead to an “overconfidence bias” that induces people to underestimate risks associated with behavior they value (Loewenstein et al. 2001).

But an alternative explanation, one in keeping with the cultural evaluator theory, is that individuals’ decisions to forgo or forbear risks is based not on the expected utility of those actions but on their social meanings, which are unlikely to be tied in any systematic way to the actuarial magnitude of those risks. The individualist, for example, who continues to worry more about being rendered defenseless than about being shot as the risks of insufficient gun control appear to increase might “not so much [be] afraid of dying as afraid of death without honor” (Douglas and Wildavsky 1982, p. 6). Similarly, for the person who values an activity – say, smoking – precisely because she subscribes to an ethic that prizes the “authenticity of impulse and risk,” a cultivated disposition to discount the likelihood of personal harm may be integral to the very form of life that activity helps her to experience (Gusfield 1993). For such persons, moreover, the very idea of conforming their attitudes toward a risk

to the results of a cost-benefit calculus might bear a meaning that denigrates their values (Ackerman and Heinzerling 2004).

Another feature of popular risk perceptions that is thought to reflect the biasing effect of emotion is the tendency of individuals' assessments of risks and benefits to be inversely correlated (Alhakami and Slovic 1994). Rather than attend to information about a putatively dangerous activity in a deliberate and systematic fashion, it is said, individuals conform their assessments of all manner of information to their emotional appraisals, perhaps to avoid dissonance (Loewenstein et al. 2001). This is a plausible reading of the results of these studies. But so is the conclusion that individuals are forming (or, just as likely, reporting) the perceptions of *both* risks and benefits that best express their cultural evaluations of an activity. In that case, the inverse correlation between risks and benefits would reflect the *expressively rational* effect of cultural worldviews, and not the irrational impact of emotion, on information processing.

Another supposed sign of the influence of emotion on information processing is the responsiveness of individual risk perceptions to the *vividness* of information (Loewenstein et al. 2001). The irrational weigher theory treats this as further evidence that emotions warp reasoned analysis. Emotionally gripping depictions of harm (e.g., news coverage of a terrorism attack), it is said, are more salient than emotionally sterile ones (e.g., stories about the consequences of global warming). Accordingly, they are more likely to be noticed and recalled, generating the distorted estimation of risks associated with the "availability effect" (Sunstein 2007).

But again the cultural evaluator model offers an alternative explanation that fits the data just as well, if not better. The impact of vivid information on risk perceptions is conditional on individuals' cultural worldviews. Shown news of a school shooting spree, egalitarians and communitarians fix on the horrifying image of dead children and revise upward their assessment of the risks of private gun ownership. What captures the attention of hierarchical and individualistic persons, however, is the tragic inability of school personnel to cut the massacre short because they were forbidden by law to bring their own guns onto school premises – a dreaded outcome that causes them to revise upward their assessment of the risk of *gun control* (Kahan and Braman 2003). Likewise, terrorism risks loom larger than global warming risks *only* in the imagination of hierarchs, not in the imagination of egalitarians – and in the mind of individualists, neither is particularly worrisome (Kahan et al. 2007b). Because *all* persons of all cultural persuasions have a stake in forming an evaluation of the incident that appropriately expresses their values, there's no reason to view anyone's response to the vividness of the story as biased rather than rationally informed by emotion.

A similar conclusion can be drawn about one last feature of risk perceptions often presented as evidence of the biasing effect of emotion. This is the tendency of public risk perceptions to reinforce and feed on themselves. Irrational weigher theorists depict this phenomenon as a form of "hysteria" or "mass panic" (Kuran and Sunstein 1998). They link it to emotion by identifying the cause as "highly vivid cases . . . that receive concentrated media attention" resulting in a distorting

“interplay between anxiety, fear, and subjective probabilities” (Loewenstein et al. 2001, p. 279).

The problem with this argument is that the power of social influence to amplify perceptions of risk is also known to be highly conditional on individuals’ cultural orientations. The view that nuclear power is dangerous and that global warming is a serious threat is uniformly held by egalitarians, but almost uniformly rejected by hierarchs and individualists. Hierarchs have formed a perception that abortion is hazardous for women, but other groups have not. Egalitarians and communitarians aren’t worried that restrictions on firearms will increase the risk that violent criminals will engage in predation, but individualists are up in arms about it (as it were).

For the cultural evaluator theory, the culture-specificity of self-reinforcing risk perceptions is easy to explain. Individuals have a stake – a perfectly rational one, as people who care about meanings and not just about consequences – to form positions on risk that express their cultural values. That by itself generates a certain tendency toward uniformity of risk perceptions within groups of culturally like-minded persons. But insofar as one of the primary sources of information people have about the relationship between their values and a putatively dangerous activity is what persons who share their commitments think about it (Cohen 2003), perceptions of danger naturally feed on one another among persons who share cultural commitments (Braman et al. 2005). This form of group polarization in risk perceptions, then, is another dynamic that can be explained consistently with the view that emotion is a form of expressive perception and not a cognitive bias.

3.3 Emotion and Systematic Reasoning: Substitutes or Complements?

The experiments I have examined to this point show that emotion matters for risk perception, but they don’t address whether emotion is functioning as bias or as a form of expressive perception. A third type arguably does both.

This research relates to how information and emotion interact. The irrational weigher theory treats emotion as an heuristic, System 1 substitute for more considered, System 2 information processing. It follows from this that the situation in which a person is likely to rely *most decisively* on emotion is when she must form an instantaneous judgment about a risk about which she has little or no information. As people obtain more information and have more time to reflect about a novel risk, their judgments should be less affective or emotional. In this sense, then, the irrational weigher theory hypothesizes a negative interaction between information and emotion.

The cultural evaluator theory suggests something different. According to that theory, emotion enables a person to form an attitude about risk that appropriately expresses her values. Emotion can’t reliably perform that function, however, if a person lacks sufficient information to form a coherent judgment about whether crediting it would affirm or denigrate her worldview. On this account, then, emotion can

be expected to play a *bigger* role in the judgment of someone who has had access to information and time to reflect on a relatively novel risk than someone who has not. In this sense, the cultural evaluator theory predicts a positive interaction between information and emotional perception of risk (Fig. 4).

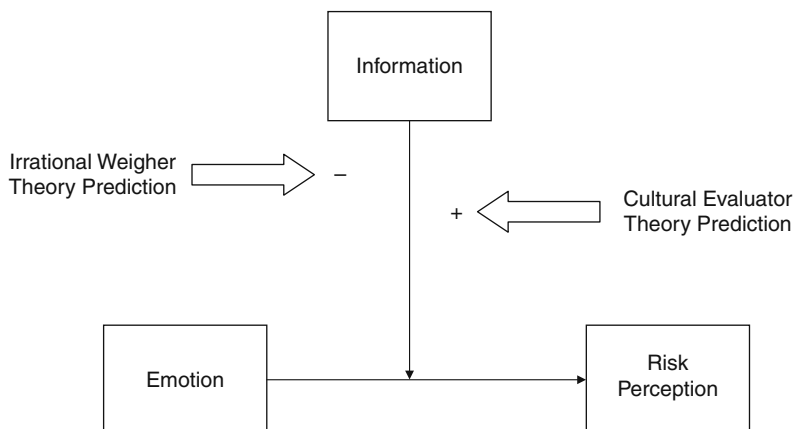


Fig. 4 Hypothesized interactions of information and emotion

Paul Slovic, Don Braman, Geoff Cohen, John Gastil, and I conducted an experiment to test these competing hypotheses (Kahan et al. 2007c). We assessed peoples' perceptions of the risks of nanotechnology. As we expected, the vast majority of our subjects – about 80% – had heard either “little” or “nothing” about this technology before we conducted our study. Nevertheless, close to 90% had an opinion on whether nanotechnology's potential risks would outweigh its potential benefits. Not surprisingly, their affective responses to nanotechnology exerted a strong influence on their perceptions. But consistent with the prediction of the cultural evaluator theory, and inconsistent with that of the irrational weigher theory, the impact of affect relative to other influences (such as gender, race, or ideology) was significantly *larger* among persons who knew a modest or substantial amount about nanotechnology before the study. Likewise, we found that affect, as well as cultural worldviews, played an even bigger role in explaining variation among subjects who received information about nanotechnology before their views were elicited than in those who did not receive information first. Again, these findings suggest that emotion is not a heuristic substitute for information, but rather a type of evaluative judgment that depends on access to enough information for a person to evaluate the social meaning of a putatively dangerous activity (Fig. 5).

Is this study conclusive in the contest between “emotion as bias” and “emotion as expressive perception”? Definitely not. But as the only study that puts the two squarely in conflict, it underscores the importance of resisting the fallacious inference that because emotion does not perform the role assigned to it by the (discredited) rational weigher model, the function it performs must be an irrational one.

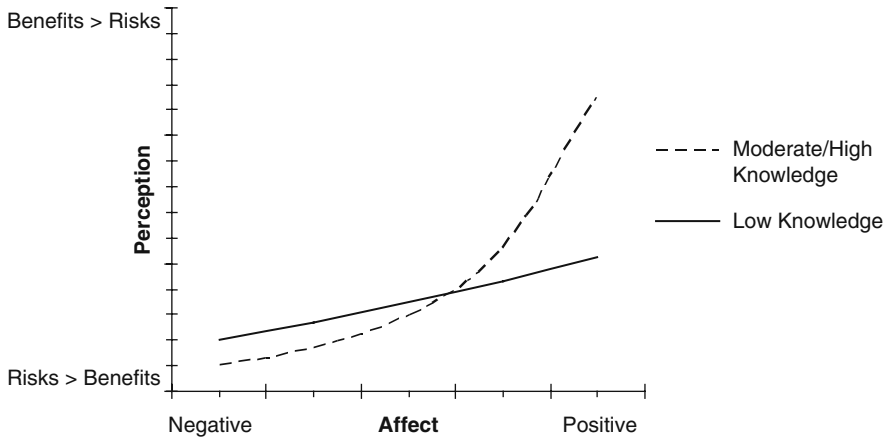


Fig. 5 Differential impact of affect on nanotechnology risk perceptions based on prior knowledge

4 Normative and Prescriptive Implications

Only the conceptions of emotion associated with the irrational weigher theory and the cultural evaluator theory fit the data on the relationship between emotion and risk perception. I now want to consider what is at stake as a practical matter in the conflict between them. Whether we see emotion as bias or expressive perception, I will argue, has immense normative and prescriptive implications for risk regulation.

4.1 Expertise—Scientific and Moral

The normative program associated with irrational weigher theory has two adversaries. One is a largely anti-interventionist stance that counsels that market forces be trusted to set appropriate levels of risk absent manifest externalities, which themselves should be remedied through regulations that “mimic” the risk-benefit tradeoffs reflected in well-functioning markets (Gillroy 1999; Viscusi 1983). If, as the irrational weigher theory asserts, emotions pervade and distort popular beliefs about risk, then there is little reason to assume that the decisions people make about their own welfare furnish a reliable guide for regulation (Akerlof and Dickens 1982). The other adversary is a fundamentally “populist” regime that favors reliance on highly participatory democratic processes to identify appropriate levels of risk. That strategy, according to irrational weighers, assures convulsive regulatory responsiveness to the alternating currents of myopia and hysteria that animate popular risk perceptions (Breyer 1993; Sunstein 2005).

In place of these approaches, the irrational weigher theory advocates delegation of regulatory authority to politically insulated, scientifically trained risk experts. These individuals, it is said, have the information and technical acumen necessary

to engage in reflective, System 2 reasoning, free of the biasing effects of emotion. By installing experts in independent regulatory agencies with which politicians cannot (easily) interfere and to which courts are obliged to defer, the law inoculates them from the virus of public irrationality.

Contrary to the objections of the defenders of the pro-market and populist strategies, moreover, irrational weigher theorists argue that this essentially depoliticized mechanism for intervening in private decision making need not be viewed as disrespectful of either individual freedom or self-government. Since ordinary people presumably would disown beliefs that are the product of emotional irrationality, regulating them via standards set by independent experts instead conforms their conduct to the preferences they would hold, as individuals and as a society, if they had the cognitive capacity to form considered and rational beliefs. "When people's fears lead them in the wrong directions," Sunstein explains, this form of "libertarian paternalism can provide a valuable corrective" (Sunstein 2005, p. 7).

The cultural evaluator theory suggests a strong critique of this defense of virtual-representation-by-risk-expert. According to the cultural evaluator model, most of the phenomena that the irrational weigher theory attributes to emotionally biased decision making in fact reflects the use of emotion to form expressively rational stances toward risk. If individuals' factual beliefs *are* expressive of cultural worldviews, then experts who treat those beliefs as "blunders" unentitled to normative respect in a "deliberative democracy" (Sunstein 2005, p. 126) are necessarily shielding regulatory law from citizens' visions of the good society. In fact it is quite debatable whether risk experts' judgments are as impervious to emotion as irrational weigher theorists believe (Slovic 2000). But however much more they know than ordinary members of the public about the actuarial magnitudes of various risks, the scientific experts certainly possess no special insight on the cultural values society's laws should express (Douglas and Wildavsky 1982).

It is exactly this mismatch between the sort of technical expertise possessed by risk experts and the *emotional* expertise needed to connect stances toward risk to citizens' values that informs unease toward "cost-benefit" and related welfarist modes of policymaking (Ackerman and Heinzerling 2004). It's not impossible to imagine the law being coherently informed by such methods. What *is* impossible to imagine, though, is that the policies will adequately engage the difficult expressive questions that risk conflicts inevitably present. If part of what's troubling (to some) about nuclear power is what it would *say* about our values to leave to future generations the problem of dealing with ever-accumulating and forever-toxic wastes, then how does it help to treat the likelihood that future generations will in fact find a solution as just another variable in the cost-benefit calculus? If part of what disturbs (some) people about gun control is the condition of servility it expresses to cede protection of themselves and their families exclusively to the state, how responsive is it to print out a regression analysis that shows more lives are saved on net than are lost when hand guns are banned? A form of policymaking that deliberately *excluded* the expressive insight uniquely associated with emotional perception would leave a society in a morally disabled posture analogous to the state of impairment experienced by the emotion-free individuals Damasio describes (Damasio 1994).

Nevertheless, this objection to deferring to scientific risk experts does not commit the cultural evaluator theory to either the pro-market or populist programs of risk regulation. Recognizing that emotions enable persons to perceive expressive value doesn't imply that the insight it imparts can never be challenged (Nussbaum 2001). Indeed, the idea that emotions express cognitive evaluations is historically conjoined to the position that emotions can and should be evaluated as true or false, right or wrong, reasonable or unreasonable, in light of the moral correctness of the values those emotions express (Kahan and Nussbaum 1996).

When we appreciate the expressive contribution that emotions make to risk perception, we are equipped to discern issues of justice that never come into focus under welfarist styles of risk assessment. Should a person about to be operated on be entitled to information about the risk that he could contract HIV from an infected surgeon (McIntosh 1996)? Why not, if we think of the decision as reflecting only the interest a prospective patient has in calculating the costs and benefits of her treatment options? But what should our answer be if we know that fear of this risk – at least in those who placidly tolerate many larger risks incident to surgery – expresses commitment to a hierarchical worldview that condemns forms of deviance symbolically associated with AIDS (Kahan et al. 2006)? Is it appropriate for a legislature to limit access to guns in order to avoid the risk of shooting accidents or violent crime? The question is at least a more complicated one if we recognize that part of what motivates aversion to these risks is an egalitarian and communitarian cultural style that despises the individualistic connotations of private gun ownership (Kahan and Braman 2003).

Analogous, and equally difficult, questions arise in other areas of law in which emotions figure (Kahan and Nussbaum 1996). No set of procedures or doctrines, in my view, can ever assure that these issues will be resolved in a just way.

But the normative complexity that the cultural evaluator theory injects into risk regulation is by no means a reason to shy away from it. For if emotion does indeed figure in our risk perceptions in the way that that theory implies, we would certainly be fools not to recognize how dependent risk regulation is on moral as well as scientific expertise.

4.2 *On Education of the Emotions*

Even if risk regulation is not *just* about promoting societal welfare measured in instrumental terms, it is still *significantly* about that. As divided as they might be in their interests in what the law *says*, hierarchists and egalitarians, individualists and communitarians surely have a common interest in what the law *does* to secure them from environmental catastrophe, from disease, from market collapse, and from attacks upon the nation's security. What do the two conceptions of emotion in risk perception imply about the prospects for making the law responsive to the best scientific knowledge we have on how to achieve these ends?

The irrational weigher theory's message is a discouraging one. Trying to educate citizens, according to proponents of this view, is even worse than futile. Not only

do citizens lack the time and capacity to engage scientifically complex data on risk in a considered, dispassionate way, but precisely because they don't, exposing them even to empirically sound information will often do more harm than good (Sunstein 2005, p. 125):

Government is unlikely to be successful if it simply emphasizes the low probability that [a feared] risk will come to fruition. The best approach may well be this: *Change the subject*. . . . [D]iscussions of low-probability risks tend to heighten public concern, even if those discussions consist largely of reassurance. Perhaps the most effective way of reducing fear of a low-probability risk is simply to discuss something else and to let time do the rest.

The cultural evaluator theory, however, generates a more optimistic conclusion. Historically, the view that emotions are "judgments of value" has also been affiliated with the position that emotions can be educated. The type of instruction this approach contemplates, however, consists not in a stoic program of disciplining the mind and strengthening the will to resist the supposedly corrupting influence of emotion on judgment. Instead, it has involved a species of *moral* instruction that reforms a person's emotional apprehension of the social meanings that unjust or destructive states of affairs and courses of action express (Nussbaum 2001).

Emotional evaluations of risk are likewise subject to education. As the nanotechnology study shows, individuals' emotions *are* responsive to information. What individuals' emotions respond to as they learn more, however, is not the expected utility of forgoing or forbearing particular risks, but rather the social meaning of doing so. The prospects for making members of the public receptive to sound empirical information, then, doesn't depend on whether they can be trained *not* to apprehend risk through their emotions; it depends on whether scientifically sound information can be made to bear a social meaning that fits citizens' cultural values.

As I have discussed elsewhere, (Kahan et al. 2006) the cultural evaluator theory suggests that this objective can be achieved through a risk-communication strategy that employs *cultural identity affirmation* and *expressive overdetermination*. In effect, individuals are cognitively motivated to reject information about risk when they perceive that accepting it would threaten their defining group commitments. To avoid this reaction, then, information about risks must be framed in a way that *affirms* rather than denigrates recipients' cultural identities; to make it possible for persons of diverse cultural persuasions to experience that affirmation simultaneously – and thus reach consensus on a contested risk issue – the information must be framed in a way that expresses a *plurality* of social meanings.

There are many examples of this type strategy in action. The adoption of tradable emissions – a market mechanism for controlling pollution – made it possible for individualists, hierarchists, egalitarians and communitarians to accept information about effective policies for securing clean air. The proposal to use nuclear power to reduce reliance on fossil fuel energy sources responsible for global warming is making hierarchists and individualists more receptive to information about the seriousness of climate change and egalitarians and communitarians more receptive to

information about the feasibility of safely producing nuclear energy. Donald Braman and I have proposed policies that use identity affirmation and expressive overdetermination to help contending cultural groups converge on sound information about gun risks (Braman and Kahan 2006).

Whether a program of “deliberative risk communication” of this type can succeed is admittedly an open question. But because it offers the only serious hope for making the complex task of risk regulation amenable to meaningful self-government, the risk of its failure is well worth taking.

5 Conclusion

In this essay, I have examined both the growing evidence on emotions and risk perception and how that evidence should be interpreted. It is settled at this point that emotions play a critical role in the cognition of risk, a finding that further undermines the already tenuous foundations of the classic, “rational weigher” theory of risk perception. But commentators, I have argued, have been much too quick to infer that emotions therefore contribute to the deformation of public risk perceptions asserted by the now dominant “irrational weigher” theory. Another conception of emotion – not as bias but as expressive perception – fits the evidence just as well (indeed, perhaps even better). On this account, emotions play a critical role in perfecting the function that risk perceptions play as rational expressions of value under the emerging cultural evaluator theory.

The recent literature on the role of emotion in risk perceptions, then, has not resolved the classic debate on the relationship between emotion and reason. It has only moved that debate to a new location, one in which the stakes are incredibly high. An error in one direction could compromise our society’s safety and welfare. But an error in the other could just as easily cost the public a meaningful voice in deciding how our society should address the major issues of our time.

We should proceed with an open mind in our continued investigation of what emotion contributes to risk perception and what its significance is for risk regulation. But we ought to be motivated as well by a morally discerning fear of all we stand to lose if we reach the wrong conclusion.

Acknowledgements Research for this paper was funded by The National Science Foundation (Grant SES 0621840) and by the Oscar M. Ruebhausen Fund at Yale Law School. I am grateful to Meredith Berger for editorial assistance. An earlier version of this essay was published in the *Pennsylvania Law Review* 156, 741–66 (2008).

References

- Ackerman, F., and L., Heinzerling. 2004. *Priceless: On Knowing the Price of Everything and the Value of Nothing*. New York: New Press.
- Akerlof, G. A., and W. T., Dickens. 1982. The economic consequences of cognitive dissonance. *The American Economic Review* LXXII: 307–319.

- Alhakami, A. S., and P. Slovic. 1994. A psychological-study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis* 14(6): 1085–1096.
- Anderson, E. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Berndsen, M., and J., van der Pligt. 2005. Risks of meat: The relative impact of cognitive, affective and moral concerns. *Appetite* 44: 195–205.
- Braman, D., D. M., Kahan, and J., Grimmelmann. 2005. Modeling facts, culture, and cognition in the gun debate. *Social Justice Research* 18: 283–304.
- Braman, D. K., and M., Dan. 2006. Overcoming the fear of guns, the fear of gun control, and the fear of cultural politics: Constructing a better gun debate. *Emory Law Journal* 55(569): 588–595.
- Breyer, S. G. 1993. *Breaking the Vicious Circle: Toward Effective Risk Regulation*. Cambridge, MA: Harvard University Press.
- Cohen, G. L. 2003. Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology* 85(5): 808–822.
- Curtis, V., and A., Biran. 2001. Dirt, disgust, and disease: Is hygiene in our genes? *Perspectives in Biology and Medicine* 44(1): 17–31.
- Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
- Douglas, M. 1966. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. New York: Frederick A. Praeger.
- Douglas, M., and A. B., Wildavsky. 1982. *Risk and Culture : An Essay on the Selection of Technical and Environmental Dangers*. Berkeley: University of California Press.
- Elster, J. 1999. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.
- Fischhoff, B., P., Slovic, and S., Lichtenstein. 1977. Knowing with certainty: The appropriateness of extreme overconfidence. *Journal of Experimental Psychology: Human Perception and Performance* 3: 552–564.
- Frederick, S. 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* 19(4): 25–42.
- Gillroy, J. M. 1999. Environmental risk and the traditional sector approach: Market efficiency at the core of environmental law? *Risk Analysis* 10: 139, 145.
- Gusfield, J. R. 1993. The social symbolism of smoking and health. In *Smoking Policy: Law, Politics, and Culture*. R. L. Rabin, and S. D. Sugarman, eds., 49, New York: Oxford University Press.
- Jolls, C., C. R., Sunstein, and R., Thaler. 1998. A behavioral approach to law & economics. *Stanford Law Review* 50: 1471.
- Kahan, D. M., and D., Braman. 2003. More statistics, less persuasion: A cultural theory of gun-risk perceptions. *Pennsylvania Law Review* 151: 1291–1327.
- Kahan, D. M., and D., Braman. 2006. Cultural cognition of public policy. *Yale Journal of Law and Public Policy* 24: 147–170.
- Kahan, D. M., D., Braman, J., Gastil, P., Slovic, and C. K., Mertz. 2007a. Culture and identity-protective cognition: Explaining the white-male effect in risk perception. *Journal of Empirical Legal Studies* 4(3): 465–505.
- Kahan, D. M., D., Braman, P., Slovic, J., Gastil, and G. L., Cohen (2007b). The Second National Risk and Culture Study: Making Sense of – and Making Progress In – The American Culture War of Fact, from <http://ssrn.com/paper=1017189>
- Kahan, D. M., and M. C., Nussbaum. 1996. Two conceptions of emotion in criminal law. *Columbia Law Review* 96: 269.
- Kahan, D. M., P., Slovic, D., Braman, and J., Gastil. 2006. Fear of democracy: A cultural critique of Sunstein on risk. *Harvard Law Review* 119: 1071–1109.
- Kahan, D. M., P., Slovic, D., Braman, J., Gastil, and G. L., Cohen (2007c). *Affect, Values, and Nanotechnology Risk Perceptions: An Experimental Investigation*: SSRN.

- Kahneman, D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review* 93(5): 1449–1475.
- Kahneman, D. et al. 1991. The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives* 5: 193, 197–199.
- Kuran, T., and C. R., Sunstein. 1998. Availability cascades and risk regulation. *Stanford Law Review* 51: 683.
- Lessig, L. 1995. The regulation of social meaning. *The University of Chicago Law Review* 62: 943–1045.
- Loewenstein, G. F., E. U., Weber, C. K., Hsee, and N., Welch. 2001. Risk as feelings. *Psychological Bulletin* 127(2): 267–287.
- McIntosh, P. L. 1996. When the surgeon has HIV: What to tell patients about the risk of exposure and the risk of transmission. *University of Kansas Law Review* 44: 315.
- Nussbaum, M. C. 2001. *Upheavals of Thought: The Intelligence of the Emotions*. Cambridge, NY: Cambridge University Press.
- Peters, E., and P., Slovic. 2007. Affective asynchrony and the measurement of the affective attitude component. *Cognition & Emotion* 21(2): 300–329.
- Philipson, T. J., and R. A., Posner. 1993. *Private Choices and Public Health: The AIDS Epidemic in Economic Perspective*. Cambridge, MA: Harvard University Press.
- Posner, R. A. 2004. *Catastrophe: Risk and Response*. New York: Oxford University Press.
- Posner, R. A. 2006. *Not a Suicide Pact: The Constitution in a Time of National Emergency*. New York: Oxford University Press.
- Siegrist, M., T. C., Earle, H., Gutscher, and C., Keller. 2005. Perception of mobile phone and base station risks. *Risk Analysis* 25(5): 1253–1264.
- Slovic, P. 2000. *The Perception of Risk*. London ; Sterling, VA: Earthscan Publications.
- Slovic, P., E., Peters, M. L., Finucane, and D. G., MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24(2): 311–322.
- Stocker, M., and E., Hegeman. 1996. *Valuing Emotions*. Cambridge [England] ; New York: Cambridge University Press.
- Sunstein, C. R. 2002. Probability neglect: Emotions, worst cases, and law. *Yale Law Journal* 112(1): 61–107.
- Sunstein, C. R. 2005. *Laws of Fear: Beyond the Precautionary Principle*. Cambridge, UK; New York: Cambridge University Press.
- Sunstein, C. R. 2007. On the divergent American reactions to terrorism and climate change. *Columbia Law Review* 107(503): 507.
- Viscusi, W. K. 1983. *Risk by Choice : Regulating Health and Safety in the Workplace*. Cambridge, MA: Harvard University Press.
- Wildavsky, A., and K., Dake. 1990. Theories of risk perception: Who fears what and why? *Daedalus* 114: 41–60.

Emotions Within the Bounds of Pure Reason: Emotionality and Rationality in the Acceptance of Technological Risks

Dieter Birnbacher

1 Discrepancies in the Assessment of Technological Risks by Experts and Laypeople

One thing that the last decades of technological innovation have shown is that there is a sharp divergence between the perception, assessment and acceptance of technological risks by laypeople and by scientific and technological experts (cf. Slovic et al. 1979; Renn and Zwick 1997, 87 ff.; Renn 2008). It has become evident that the general public, in judging the acceptability of technologies, makes use of mental “heuristics” (Tversky and Kahneman 1974) that differ significantly from those employed by experts. This has become obvious in the case of energy production by nuclear power and of transgenic crop plants in agriculture, two cases that have led to well-known and enduring social and political schisms. Moreover, it has led to a deep and lingering distrust between representatives of science, technology and industry on the one side and great portions of the general public on the other. A view widespread among scientists, engineers and industrialists is that some if not even most factors underlying the public acceptance of technologies do not mirror any objective features of these technologies and their prospects but are merely “psychological” and, in the last analysis, irrational. Public attention, according to this view, is focussed on risks that hardly ever affect the life of those who fear them, whereas other, far more dangerous risks, are ignored. The point is nicely expressed by Peter Sandman’s dictum that “the risks that kill you are not necessarily the risks that anger and frighten you” (cf. Jungermann and Slovic 1993, p. 80). In this view, the fears, anxieties and worries manifested by the stubborn non-acceptance of technologies like nuclear power and gene-food express reactions that may be psychologically understandable (or at least explainable) but only thinly related to the facts. One form this criticism assumes is that these reactions, however firmly embedded in the human psyche, are “purely emotional”. From this diagnosis it is only a short step to the conclusion that it is wrong, and possibly even irresponsible, to give these emotions a role to play in technological planning and development.

D. Birnbacher (✉)
Department of Philosophy, Düsseldorf University, Düsseldorf, Germany
e-mail: Dieter.Birnbacher@uni-duesseldorf.de

Accordingly, emotions are a factor politicians have to deal with, not technological planners. Emotions are a matter of strategy, not something that should enter into judgements about what is acceptable or unacceptable.

A similar discrepancy can be found on the level of theory. The scientific debate about acceptable risk largely recapitulates the public debate, with a rational choice approach on the line of a socially extended Bayesianism on the one side and a theory of “qualitative characteristics” (Slovic et al. 1979, p. 36) on the other. On the Bayesian side, risks are assessed from a risk-neutral perspective, and only the two “classical” dimensions of risk are taken into account, (negative) value of possible outcomes and outcome probability. On the other side, a risk-averse perspective is held to be more appropriate and a number of further aspects are brought into the picture, among them voluntariness, potential for catastrophe, naturalness and distributional aspects. There are obvious advantages in the Bayesian approach from a practical point of view. It makes the calculation and comparison of risks considerably more straightforward than an approach that takes into account a large number of (probably weighted) qualitative factors. Whereas in Bayesianism comparisons between the benefits and risks of technological alternatives involve comparing one single value, expected utility, risk assessments by qualitative standards are inherently more complicated, intransparent and controversial. Whereas expected values are derived by a simple arithmetic operation (the summing of the products of each individual outcome with its probability), risk assessments by qualitative factors usually fail to make explicit which factors are included and what relative weight is given to each of them, thus lowering the chances of a systematic, authoritative and consensual assessment. The moot question, however, is how these more or less formal advantages of Bayesianism compare with its material inadequacies and especially its inability to pay tribute to some of the factors paramount in the acceptance or non-acceptance of a technology by the general public. One of these inadequacies is that Bayesianism assesses risky activities by a purely additive criterion and is indifferent between risks with great magnitude and low probability and risks with low magnitude and high probability. That means that it is unable to account for at least one of the structural asymmetries that are relevant in risk acceptance. For the general public, it matters a lot whether the risk profile of a technology includes severe but infrequent accidents or frequent but trivial accidents. And since Bayesianism balances costs with benefits irrespective of the identities of their subjects it cannot take account of distributional features. It cannot distinguish between cases in which the harm resulting from a technology befalls those who benefit from it and others in which it is imposed on third parties.

The conflict between these viewpoints has persisted for decades, and with a high degree of polarization. Both parties insist on the rationality of their respective perspective and do not hesitate to accuse the other of irrationality, blindness and yielding to “pure emotion”, often, ironically, with a good deal of emotion. On the Non-Bayesian side, the same “emotional” factors in risk perception and assessment – such as the preference for the natural and the familiar against the technical and unfamiliar – are interpreted as indicators of intuitive wisdom that are denounced on the Bayesian side as sentimentality and ignorance. The tendency to substitute

unanalysed intuition for systematic decision-theoretical approaches is particularly prominent in some variants of “bounded rationality”. Gigerenzer and Selten, for one, have pleaded for an approach to complex choices that no longer attempts to integrate qualitative factors into the Bayesian scheme by “tinkering with the utility or probability function, while at the same time retaining the ideal of maximization or optimization” (Gigerenzer and Selten 2001, p. 3). Instead, they recommend to question the monopolistic claims of Bayesianism and to adopt a radically new approach that starts from the heuristics factually employed in choices, mostly on an unconscious and “instinctive” basis, the “adaptive toolbox”. In this way, decisions are not only more readily available but also better adapted to their respective contexts. The interesting point for our purposes is that, according to these authors, some of the emotions share these advantages. There are circumstances in which emotions are a more reliable guide to rationality than deliberation and calculation, as is shown in the context of animal behaviour:

Emotions like disgust or parental love can provide effective stopping rules for search and a means for limiting search spaces. In particular, for important adaptive problems such as food avoidance, an emotion of disgust, which may be acquired through observation of conspecifics, can be more effective than cognitive decision making. (Gigerenzer and Selten 2001, p. 9)

This plea for “relying on instinct” will, however, make no impression on the Bayesian who will be quick to point out that these authors measure the success of the method by an independent standard of rationality (“effectiveness”) that is not supposed to be determined by emotional factors. Furthermore, Bayesians will be ready to concede that in many real-life contexts emotions may indeed be more reliable guides to behaviour than deliberation and calculation, but that this does in no way detract from the necessity to establish a standard of rationality on which the instrumentality of “rules of thumb”, “gut-feelings”, “instincts” and the like can be judged. The standard invoked in judging whether a “tool” from the “adaptive toolbox” is truly “adaptive” (instead of “maladaptive”) cannot itself be justified with reference to the tools from the box. After all, Gigerenzer and Selten are far from denying that there are plenty of occasions in which the ready-made “tools” commonly employed in practical matters are of little help, or even positively misleading.

The Bayesian, however, will go further and maintain that there is evidence from a great number of empirical studies that the effects of emotional factors in judgements on risks are predominantly of a distorting kind. This evidence is particularly strong since it does not have to rely on some presupposed standard of rationality but points to features that are incompatible with any coherent standard:

1. There is strong evidence that risk perception and risk assessment are to a high degree culture-relative and depend on non-cognitive factors like habituation and dissonance reduction. The most plausible explanation for the differences in the judgements about the risks of nuclear energy between the populations of France and Germany, or in the judgements about transgenic food between the populations of Europe and the United States is not that one of these publics is better informed than the other or in a better position to draw the right conclusions

from this information, but that judgements are harmonized with what has become part of their habitual environment. The basis of this harmonization of facts and values is not coherence in the sense of a unified cognitive picture of the world but coherence in the sense of dissonance reduction, the unconscious striving for a view of the world in which action, cognition and emotion are in harmony with each other.

2. There is evidence that there are sharp discontinuities in popular risk perception. There seems to be a threshold in probability (sometimes identified with the probability 10^{-5}) beneath which rare events are judged to be sufficiently improbable to be no longer a cause of concern. We do not care about catastrophes that are “morally (or practically) impossible”. However, we care a lot about catastrophic risks the probabilities of which are only slightly higher. Whereas risks below the threshold are discounted and ignored, risks above the threshold are judged as particularly dangerous. On the theoretical level, this dichotomy is mirrored in Nicholas Rescher’s theory of acceptable risk that makes a similarly sharp contrast between risks of catastrophe too trivial to be given any attention, and non-trivial risks of catastrophe that must be avoided at all costs, thus making the overall negative value of risks leap from zero to infinite at the threshold (cf. Rescher 1983, p. 76).

What is blurred by the polarization of Bayesians and Non-Bayesians is the prospect of finding a compromise that does justice, as far as it goes, to both sides and attempts to combine what is adequate in both perspectives. This is what I propose to do in the following. The question, in my view, is not whether the one or the other approach is the correct one, but how far the discrepancy between them is real or apparent, and how far the Bayesian model can do justice to the attitudes behind the qualitative risk features, provided these can be shown to encapsulate aspects of rationality not covered by the standard variants of Bayesianism. Before looking at this question, we should, however, first clarify in what exact way emotions can be said to influence judgements about risks and how far they are acceptable.

2 The Role of Emotion in Judgements About Acceptable Risk

“Emotion” serves as an umbrella concept for a large variety of mental items, and in discussing the role of emotion in judgements about risks one should make clear what kind of item one has in mind. The question as to what extent emotions can serve as avenues to the truth where reason is blind, and to what extent they distort and mislead judgement, depends, among others, on what category of emotion one is thinking of. There are two categories of emotion for which it is more or less obvious that their influence on judgement is mainly a distorting one and that they justify, so far as it goes, the view of philosophers like the Stoics for whom emotions were, in the first place, “disturbances” – not only in the sense that emotions disturb one’s peace of mind but also in the sense that they disturb one’s sound judgement. These

two categories are emotions as *temporally extended moods* and emotions as *episodic forms of excited feeling*.

It is a fact well-known from experience that both kinds of emotional states tend to weaken our faculty of judgement and to engender pessimism or optimism, as the case may be, largely uncontrolled to the facts. In a depressive mood, things appear threatening that are normally seen as indifferent or easy to cope with. In an elevated mood, the “bright side of life” dominates inner experience, shielding from view what might blemish the harmonious picture. Similar observations can be made in cases of acute emotion. In an experiment on the impact of affect on risk assessment Johnson and Tversky investigated to what extent fear, anxiety and worry caused by the reading of dramatic stories have an influence on estimates of the frequencies with which certain risks are believed to occur. The subjects were made to read stories with vivid and detailed portrayals of deaths deliberately designed to induce anxiety and worry. It turned out that there was a considerable influence of the emotional states induced by the stories on the estimates of the frequency of certain risks though the risks had nothing to do with the dramatic events in the stories. The frequency estimates of the group with induced negative mood was significantly higher than those of the control group. An analogous experiment with a group with induced positive mood showed an even higher inverse effect (Johnson and Tversky 1983, p. 28).

These two categories of emotions, then, can be left to themselves. They are not what is at stake in the debate about the rationality and irrationality of emotion in risk assessment. The kind of emotion that is at stake can be characterised by the following features:

1. The emotions that are candidates for being honoured in judgements about acceptable risk are of the nature of emotional attitudes rather than of the nature of episodic emotions. Ontologically, they are dispositions rather than events of processes. “Emotional attitude” means that the attitude is not purely cognitive and that its content cannot be adequately expressed by a purely descriptive statement. In this sense, a belief that something is the case is purely cognitive, but not the hope or fear that something is the case. Emotional attitudes necessarily include an element of evaluation, positive or negative.
2. The emotional attitudes in question are intentionally related to the risk in question. They are closely related to the thought of the risk and are not, as the emotional states in the experiments of Johnson and Tversky, induced by independent factors.
3. The emotional content of the attitude is largely, or wholly, unconscious or pre-conscious. The subject is not aware of this emotional content, or is made aware of this content only by directing attention at it.
4. Emotional content can enter into judgements about risks at several different stages and influence components of these judgements to different degrees. Some emotional factors work primarily on the estimate of frequencies, others primarily on the estimate of the negative values of risks. One mechanism involving emotional factors and primarily influencing frequency estimates is the “availability

heuristic” (Slovic et al. 1979, p. 15). It makes that a certain event is judged as likely or frequent to the degree that it is easy to imagine or recall, which in turn depends, among others, on emotional factors like surprise, irritation and dread. Other mechanisms involving emotional factors work on the valuation component, such as the mechanism that makes risks that are a threat to ourselves look more frightening than risks that are a threat to other people.

It is exactly these emotional factors in the attitudes towards risky technologies that the dispute between Bayesians and Non-Bayesians is about. Both differ radically in their views about the compatibility of these factors with rationality as a standard of judgement and guide to action. It should not, however, be overlooked that there is also a great area of agreement between the parties as far as the compatibility and incompatibility of these factors with rationality is concerned. This agreement pertains primarily to components of judgements about risks that are consequences of more general features and independent of the probabilistic nature of the items judged.

Neither party denies that at least one non-cognitive factor is not only compatible with rational judgements about acceptable risk but even required by them, namely the non-cognitive factors entering in the estimates of the moral and non-moral value of the possible adverse events and their consequences. Statements about risks (like statements about chances) are, as a rule, value judgements and not purely descriptive. They imply that certain possible outcomes are judged to be of negative value. If, however, value judgements, as I think they do, necessarily contain a non-cognitive element and express a “pro-” or “con-attitude” with at least a minimum of emotional content, emotional attitudes are, as it were, interwoven with risk judgements and inseparable from them. By necessarily referring to values of some kind or other, the very concept of risk seems unintelligible if defined in a purely descriptive way. Risk judgements are inherently value judgements and go beyond what can be attained by a purely cognitive approach.

This leaves open the possibility that the value judgements contained in judgement about risk (and therefore about acceptable risk) may be influenced by more specific emotional factors that make them “irrational” in one of various ways. The most important of these factors are *involvement* or *ego-preference*, the tendency to judge risks to be unacceptable in relation to the extent one is threatened by them in one’s own person, and the tendency to discount adverse events according to social distance and to distance in time. I will not here discuss if there are circumstances under which these tendencies, which play an important role in common-sense judgements about risks, can be given a rational justification (cf. Birnbacher 2003). Let it suffice to say that if these tendencies are criticized as “irrational” this is not because this follows from the fact that they are non-cognitive or emotional tendencies, but because they constitute specific emotional tendencies that tend to distort judgements about the acceptability (from a moral and impersonal perspective) of actions and strategies not only in the domain of risks but likewise in non-probabilistic domains.

Furthermore, both parties are agreed that there are a number of non-cognitive factors that are incompatible with the rationality of judgements about acceptable risk but which notoriously enter into such judgements. One such very general factor is the *framing effect* explored by Kahneman and Tversky that tends to determine reactions to risks by the way risk statements are formulated. The risks of an operation tend to be judged as more acceptable if they are expressed in positive terms, i. e. in terms of the probability of survival, than if the same risks are given a negative wording in terms of the probability of death. The striking thing about this effect is that it works even with people who think they are intelligent and critical enough to be immune to verbal deception. Again, this effect does not depend on the probabilistic nature of the events in question. The framing effect seems to be a general phenomenon of communication and not specific to communication about chances and risks (cf. Tversky and Kahneman 1981, p. 457).

Another point on which both parties are agreed is that there are emotional attitudes to risks that are “irrational” in so far as they are incompatible with the considered judgements of the judging person himself. Thus, a person may develop a generalized fear of dogs after having been severely bitten by a dog that co-exists with the belief that most dogs are innocuous, and even with the knowledge that this particular fear is neurotic and unfounded. This shows that emotions and emotional attitudes can be “irrational” in more than one way. They can be “irrational” or inadequate, as this last example shows, by becoming autonomous, dissolving the ties that normally bind them to the faculty of judgement. And they can be “irrational” or inadequate by being in harmony with the faculty of judgement, which in turn is misled, either by the impact of the emotional factors involved or by independent factors. It is one of the weaknesses of most traditional theories of emotions that they do not distinguish clearly between these two kinds of irrationality. This holds even of Spinoza’s theory of emotions which in other respects goes a long way to do justice to the cognitive components of emotion and to distinguish between adequate (“active”) and inadequate (“passive”) emotions, but which nevertheless tends to regard all inadequate or “irrational” emotions as indistinguishably pathological (“delirii species”, Ethics, IV, 44 Scholium). But, of course, there is an important difference between the kind of irrationality involved in phobias and that involved in fears based on deficient judgement. If A has an “irrational” fear of a certain dog because he has a generalized phobia in respect of dogs, this can safely be categorized as pathological. This is definitely not the case if B has an equally irrational fear of a certain dog because he has been misinformed about the dog’s dangerousness or because he mistakes the dog for another dog that is in fact dangerous.

Correspondingly, there are at least two ways in which emotions and emotional attitudes are open to criticism: They can be criticised because they are neurotic, and they can be criticised because they are based on false judgments. This latter possibility is open because emotions and emotional attitudes, in contrast to moods, feelings and feeling dispositions, have judgemental components. In the case of risks, these components include, among others, judgements about the kind of consequences to be expected from a certain action or event, the values of these consequences, their frequency, and the degree of certainty associated with the estimates

of each of these dimensions. Some of these judgemental components in turn contain emotional or non-cognitive elements, such as the valuation of possible outcomes. If these judgemental components go wrong at a certain point, the emotion or emotional attitude based on it will, as a rule, go wrong as well.

3 Which “Qualitative Risk Factors” can be Integrated into the Bayesian Scheme?

There remain a number of factors about which Bayesians and Non-Bayesians differ, and these differences must now be considered. We should be careful to make two kinds of distinction from the start. The first distinction is that between those factors in judgements about acceptable risk that are in principle amenable to an analysis and evaluation within the Bayesian scheme, but are only rarely given attention in practice, and those that resist integration and require a revision of that scheme. If it turns out that the qualitative risk factors which the Non-Bayesians appeal to lend themselves to a reconstruction within this scheme, then what deserves to be criticized is the practice rather than the theory of Bayesianism. The upshot is not that thinking about technological risk in terms of values and frequencies is misguided in principle, but that the potential of this thinking is only insufficiently made use of in practice. The case is different if it turns out that some of the qualitative factors cannot in principle be reconciled with this approach.

The second distinction is that between factors in judgements about acceptable risks for which it is at least *prima facie* plausible that they should be included in a theory of acceptable risk and others for which this is less clear. We have already referred to the empirical fact that the riskiness of technologies is to some degree influenced by the estimated extent to which one is threatened by a certain risk in one's own person. It is clear that a person-relative criterion of this kind cannot legitimately figure in impersonal judgements about risks. The fact that certain consequences may be dangerous for *me* cannot be relevant to the judgement about whether a certain risky technology is acceptable from the perspective of society or of all who are positively or negatively affected by it.

Bayesianism can in principle include many of the factors that tend to be quoted by Non-Bayesians as examples of the context-sensitivity and adequacy of risk judgement of the general public. Bayesianism is an extremely flexible instrument. If many of the qualitative factors are not normally included in formal analyses this is mainly because they do not lend themselves to easy calculation. In principle, however, the costs and benefits taken account of in the Bayesian analysis are not restricted to those easy to quantify, such as money or the number of deaths or injured. Instead, they can include aesthetic, social and political benefits and harms such as the loss of amenities that go with many forms of “big” technology, the obsolescence of skills in the wake of new technologies and the loss of democratic control often involved in centralised production. On a more fundamental level, Bayesianism is not bound to one particular system of valuation of outcomes. There is no necessary connection between Bayesianism and utilitarianism, nor, for that matter,

between Bayesianism and consequentialism. Against Harsanyi who thought that the Bayesian rationality postulates “entail utilitarian ethics as a matter of mathematical necessity under relatively weak conditions” (Harsanyi 1978, p. 223) it must be said that this follows only under the presupposition that utility can be adequately conceptualized along Neumann-Morgenstern lines and that all possible values collapse into preferences. Neither is Bayesianism entailed by utilitarianism. A utilitarian can have good utilitarian reasons for a more risk-averse approach than the risk-neutral approach implied by the maximization of expected values. Though consequences matter in Bayesianism (as in all theories of acceptable risk), this does not prejudge how the (possible) consequences are valued. First, these consequences need not be valued on exclusively consequentialist terms. Even if the concept of risk is inherently a consequentialist concept in so far as it involves uncertain consequences, this does not imply that these consequences have to be assessed in accordance with an ethic that measures the severity of negativities by the value of consequences. The consequences might alternatively be measured by deontological features such as the extent to which they constitute violations of rights. As Ralph Keeney emphasized (cf. Keeney 1984, p. 120) the technical apparatus of Bayesianism is indifferent to the kind of values assigned to the possible outcomes and is able to provide even for the extreme case that the outcomes are only valued according to their moral instead of their non-moral value. In brief: Though one may easily agree to Harsanyi’s thesis that result-orientation is a central principle of rationality in responsible decision-making (Harsanyi 1978, p. 225) and that the question whether a risky activity is acceptable from a prudential or moral point of view cannot be determined by its inherent or purely symbolic features, this does not settle the issue which values, and which kinds of value, are assigned to the results. Even if it is beyond question that in decisions under risk consequences matter, this leaves open whether the standards by which the consequences are evaluated are of a consequentialist or deontological kind, and whether, if they are of a consequentialist kind, the value of the consequences is determined only by non-moral values such as life, health and quality of life or by moral values like morally good actions, morally good intentions or the exercise of virtue.

The adaptability of the Bayesian model to different systems of valuation goes even further. Apart from taking account of goods or bads that befall individuals, it is also able to incorporate structural values such as equality, equity or distributional justice provided these are operationalised in a way that makes them commensurate with individual values. Thus, it is perfectly able to incorporate the intuition that a distribution of risks is highly unfair if A gets the whole profit from a risky activity and B bears all the burdens. Empirical surveys show that judgements on acceptable risk react to his kind of unfairness (cf. Renn and Zwick 1997, p. 92). An account that aggregates only individual goods or bads cannot represent this kind of unfairness. But, as the example of Rainer Trapp’s “non-classical” brand of utilitarianism (Trapp 1988) and the “person-trade-off” approach in health economics (Nord 1999) show, these structural features can be integrated into a consequentialist scheme by treating them as a dimension of chances and risks that supplement the “classical” individualist dimensions and can be handled along the same formal lines. From an

ethical point of view, it does not at all seem incompatible with rationality to take distributional features into account. On the contrary, it seems imperative to give features like equality and fairness some role to play in the evaluation of consequences. It would be far-fetched to think that to care for equality and fairness in the distribution of risks and chances is pure sentimentality or, in this sense, purely “emotional”.

There is a further dimension of technological risk that must be taken account of in any adequate calculation and comparison of risk: the benefits of security and the harm of insecurity. The bad thing about risks is not only that they involve a bad of some kind in case they materialize. The bad thing is also the psychological threat their existence implies for those subject to the risk. The psychological benefits and harms of a technology are not exhausted by the psychological goods and bads involved in the materialization of its chances and risks. They include, in addition, the benefits and harms connected with their anticipation, especially if this undergoes a process of “social amplification” by which the psychological effects spread through society (cf. Kasperson 1988; Renn 1991). The prospect of a future possible good is, as a rule, itself a good, the prospect of a future possible bad itself a bad. Therefore, the fear and insecurity generated by the existence of a risk should be taken as serious as the feeling of insecurity generated by its materialization. In my view, they should be included in the risk profile of a technology even in cases in which these feelings seem, from an objective point of view, exaggerated or “hysterical”, or are based on severely distorted risk perceptions. As far as these feelings are immune to enlightenment they must be added to the items in the negative side of the balance, irrespective of whether they are rationally justified or not. It is significant, moreover, that according to the psychologists of risk the prospect of a future bad is worse to a higher degree than the prospect of a future good is good. Future harms, whether certain or probable, arouse more fear than future benefits, whether certain or probable, arouse joyous expectation. Obviously, we react to future harm or risk as born optimists for whom the good is the normal thing. This is an additional reason to give some weight to the feelings of insecurity generated by risks.

There is, then, a certain range of factors in laypeople’s risk perceptions and assessments that can be reconciled with Bayesianism, at least with a suitably refined version of it. It remains to be shown that this is true for all factors that can be rationally justified. Of course, as with other attempts to reconcile doctrine and common sense, we should not mislead ourselves into thinking that there is a pre-established harmony between the common sense and the scientific approach. We should take serious Amos Tversky’s warning that “in the absence of any constraints, the consequences can always be interpreted so as to satisfy the axioms” (Tversky 1975, p. 171). The best thing to keep clear of this temptation is to adopt as far as possible the point of view from which the general public judges on acceptable risk and only then make the second step to ask how this fits into the Bayesian picture.

Among the qualitative characteristics that play a role in common sense risk perception some concern primarily the *value* of the consequences of a risky activity or event, others the level of *insecurity* generated, and others both. This gives us

a principle by which we can classify the main candidates among the qualitative characteristics for integration into the Bayesian scheme.

Among the first group, the characteristic that it seems easiest to reconcile with a Bayesian approach, is *irreversibility*. Irreversible harms are commonly given more weight than reversible harms, and risks with irreversible outcomes are commonly feared and avoided to a higher extent than risks with reversible outcomes. It is evident that this factor can and must be honoured in a Bayesian analysis. In general, the fact that a harm is irreversible means that the consequences of the harm are more severe than those of a corresponding reversible harm, partly because of their scope and partly because of their opportunity costs. Irreversible harm can be expected to stay for a longer period of time than reversible harm. Moreover, it narrows the options available and in this way compromises freedom. In order to compensate for the harm in terms of subjective well-being, one usually has to invest more labour and time than in the case of reversible harm, provided that compensation is possible in the first place. Apart from material costs, psychological costs are usually higher. Whereas a house destroyed by a fire can be reconstructed, human victims cannot be revived, objects belonging to the cultural heritage cannot be restored in the original. Coping with irreversible losses of what one valued requires patience and humility, and is usually accompanied by a longer period of suffering.

Another member of the second group is (perceived) *control*. The psychology of risk has shown that risky activities that are subject to control by whoever engages in it are commonly judged to be more acceptable than comparable activities over which the subject has no control. To many people, the risks of using a ski-lift seem to be more severe than the risks of running-down skiing, the risks of travelling by airplane more severe than the risks of driving. The important variable seems to be the extent to which the subject believes to be autonomous in the direction the risky activity takes while it is running. Whereas voluntariness concerns the freedom to engage in a risky activity by one's own choice, control concerns the freedom to change the course of events at will while it lasts. Can this factor be integrated into the Bayesian scheme? Surely, at least to a certain extent. As far as control is a relevant psychological variable, it must be included in an adequate calculation of outcome values. Even if people overestimate the extent to which they are able to control the process which they think they can control, this feeling substantially contributes to the subjective feeling of safety, at least within the "Faustian" culture of the West (that successively seems to govern the world) that values active control of social and natural processes more than passive acceptance.

The most important members of the third group are *voluntariness* and *potential for catastrophe*. Both factors tend to modify both the value of possible outcomes and the extent to which risks by their very existence generate feelings of security or insecurity. *Voluntariness* has proved to be highly relevant for perceived acceptability of risks. One of the pioneers in the scientific study of risk-taking, Chauncey Starr, went so far to maintain that voluntary risks can be one thousand times as great as risks of an involuntary nature to be judged acceptable (Starr 1969, p. 1237). This conclusion, however, was derived exclusively from revealed preferences, measured in monetary terms. Though Starr's interpretation seems exaggerated, it is a

fact that people attach very great importance to having a choice instead of having risks imposed on them by others. The relevance of this factor is evident. It is evident that what matters in choices between risks is not only the estimated value of outcomes but the autonomy in taking risks in the first place. This is reflected in the striking differences in the emotional attitudes towards voluntary and involuntary risks. The harm we suffer from a self-imposed risk “feels” differently from a harm from a risk imposed by others without our consent. The risk of death by murder has an emotional quality strikingly different from that of the risk of death by suicide. Both violate our integrity, but only the former violates our autonomy. Whereas risks imposed by others or by natural factors limit our autonomy, risks imposed on ourselves by ourselves increase our autonomy. On the one hand, we have a strong interest in not being subjected to risks by others without our consent. On the other hand, we have an equally strong interest in being free to impose risks on ourselves if we so want, for example by risky kinds of pastimes and sports. Voluntariness is itself a utility, involuntariness a disutility. The attractiveness of voluntary activities lies partly in their very being voluntary, the unattractiveness of involuntary activities in their being involuntary. The same activities often assume completely different values according to whether they are freely chosen or constrained.

Voluntariness belongs to the third category because it affects both the valuation of the possible outcomes and the dimension of security. Voluntariness is not only a utility in its own right, it also has an impact on the degree to which we feel secure. Insecurity is primarily dependent on the extent to which we are subject to risks imposed by nature or by others without our free consent. It is true, the more people are prone to uncontrollable impulses the more reason they have to fear the consequences, for themselves and for others, of their own passion, rashness and foolishness. But to the same extent that they have those reasons it is doubtful whether these risks can be classified as fully voluntary.

Something similar can be said about the *potential for catastrophe*. This characteristic, too, tends to aggravate both the harm in case the risk materializes and the feeling of insecurity it generates by its existence. Risks involving harms that occur rarely but in catastrophic dimensions are much less accepted than risks with the same number of victims where these are distributed over time and each single harm is too trivial to arouse public attention. Thus, air traffic accidents are given much more publicity than car accidents though the total number of victims is considerably lower. One single accident with fifteen thousand deaths is much more spectacular than the same number of deaths by domestic accidents. But even if we discount the factor of public attention there is a strong intuitive tendency to judge risks with the potential for disaster less acceptable than risks with a more distributed pattern of incidence.

Is this intuition open to reconstruction within the Bayesian model? Certainly it is, at least to a certain extent. What distinguishes the harm caused by a catastrophic event with thousands of deaths at a time from a sequence of thousand individual deaths distributed over time is exactly that the harm occurs simultaneously and has a more thoroughgoing impact, both on the material and the psychological resources of a society. On the material side, non-linearities in the disutility of concentrated

harm have to be taken into account. One accident with the great number of victims can transcend the capacities of a society in terms of medical, technical, financial and human support. A society which can come to terms with one hundred similar incidences of disease per week is not necessarily in a position to deal adequately with five thousand incidences a day. In the long term, a catastrophic event often has lasting effects on the economy, e. g. by companies going out of business, unemployment, and costs due to rising safety standards. On the psychological side, the impact can be worse: the collapse of the economic basis of a whole region, social upheavals, loss of trust. Take as an example the impact of the Tschernobyl accident on the prestige of nuclear energy even in nations in which an imprudence similar to the one that caused the accident is hard to imagine. (To do justice to these additional effects, R. Wilson once proposed to calculate the social costs of accidents with n deaths by n^2 , cf. Starr et al. 1976, p. 657).

At the same time, the additional factor of decreased perceived security dictates that catastrophic possibilities are assigned a special weight over and above the weight they receive in expected value analysis. The very existence of the possibility that a technology can lead to catastrophic harm is a significant psychological item in the overall risk of a technology. Given the “desire for certainty” (Slovic 1978, p. 101), it makes a world of difference whether the probability of disaster is 0.00 or 0.01. This difference is much more significant than a difference between, say, a probability of 0.50 and 0.51. There are, then, excellent reasons to handle catastrophic possibilities differently from medium-sizes risks and not to level them down in the way they inevitably will be in expected value analysis as it is commonly applied. In this respect, then, the intuitive and “emotional” reactions to disastrous risks can serve as a clue. They point to the fact that the frame in which analyses are commonly carried out has to be extended so as to take account of these additional factors.

It goes without saying that paying tribute to these factors considerably complicates the Bayesian picture. The simplicity and the elegance in the evaluation of risks that recommends Bayesianism especially to engineers and technological planners would have to be sacrificed. A good deal of the sensitivity to context present in intuitive judgements of acceptable risk would have to be integrated into the Bayesian frame. But there are good reasons to justify these complications. On the one hand, any simpler version of Bayesianism would be less adequate. On the other hand, any model that renounces calculation and deals with risks on a purely intuitive basis would lack the transparency that goes with an explicit analysis and balancing of factors.

4 Leaving the Bayesian Picture Behind

One dimension that looms large in common sense risk perception and assessment has not yet been mentioned, the dimension of the perceived *uncertainty* in the probability estimates. In general, the more uncertain the probability estimate is

perceived to be, the more risk-averse is our attitude towards the risk in question. This is an important fact because with controversial technologies estimates of benefits and risks are nearly always based on limited experience and essentially depend on subjective probability estimates by experts that lack the certainty about relative frequencies available for lotteries and games.

With technologies about which there is too little experience to give reliable estimates of the frequencies with which possible harmful consequences might result, there is in fact not only one, but two kinds of uncertainty to consider, each on a different level of knowledge and ignorance: uncertainty about the probability with which certain kinds of possible adverse events are to be expected, and uncertainty about whether the list of possible consequences considered in the calculation of risks exhausts the possibilities. The first kind of uncertainty concerns, among others, risks that can be identified but cannot be calculated by standard methods like fault-tree analysis or simulation, such as common mode failures, external effects and human factor risks (cf. Kates 1981, p. 93 f.). Who, for example, would have thought of the possibility that in 1975, a technician checking for an air leak in Brown's Ferry Nuclear Plant on the Tennessee river would do this, in violation of standard operating procedures, with a lighted candle, thus causing a disastrous fire? The second kind of uncertainty is likewise hard to avoid. There is nearly always a small probability that certain risks have been overlooked or could not be known in advance, either for contingent or for principle reasons. Well-known examples from the history of technology teach us that some causes of disaster are, and can, only be identified after the event.

It can fairly be said that situations of choice with elements of uncertainty are more common than situations in which all risks are completely known. Complete knowledge of probabilities of all possible eventualities is as rare as complete uncertainty. This is especially true of situations in which technologies are at stake for which limited experience does not allow a final judgement about how safe they are under critical conditions. For these situations, the "emotional" reserve about new technologies that have not yet stood the test of proving their safety under real-world conditions, has some measure of truth in it. The generally lower acceptance of technologies with incompletely known risks (provided the chances benefits these technologies offer are not seen as substantial enough to outbalance the risks) has a *fundamentum in re* and cannot be attributed to excessive conservatism. There seems to be a "rational core" in the conservative instincts that permeate the emotional attitudes to new technologies in the general public, except in areas like medicine or communication where attention is primarily focussed on the benefits.

What does this imply for the assessment of technological risk? I think that it calls for a revision of the Bayesian model, not so much because of the inevitable uncertainty of probability estimates on the first but because of the inevitable uncertainties on the second level. There are a number of conditions that constrain Bayesianism as an appropriate strategy in risk assessment: that the risky activity or event is iterated so many times that adverse outcomes are compensated by favourable outcomes; that no outcome is so disastrous that it overthrows the system; and that all relevant benefits and risks are identified. Uncertainty on the first level can be made consonant

with these conditions by reconstructing it as a range of probability along the lines formulated, e. g., by Rescher (1983, p. 94 ff.) or by identifying a “tolerable window” (Posner 2004, p. 176 ff.). In this way, choices under uncertainty and comparisons of risk with uncertain probabilities can be treated by the same expected-value assessment appropriate to reliable probabilities. A more serious limitation is the second condition that the risks of which we are uncertain must at least admit of identification, a condition that fails to be fulfilled in many cases in which a technology is controversially discussed. This fact, indeed, is a strong argument for adopting a more risk-averse strategy than Bayesianism. This is not to say that the adequate strategy should be as conservative as the maximin rule that ranges options according to worst possible outcomes irrespective of probabilities (cf. Rescher 1983, p. 161, Leist and Schaber 1995, p. 56). Such a principle would be excessively prohibitive of technological progress, which requires a minimum of preparedness to gamble even with grave risks. But the principle should at least restrict the possibility, present in the Bayesian model, to balance severe harm for the victims of technological progress by the benefits provided to the rest of mankind.

There is an additional reason for questioning the adequacy of Bayesianism in determining acceptable technological risk. Decision-making on risky technologies can be conceived in two ways, each with different consequences for the criteria of ethical legitimacy. On the one hand, it can be conceptualised as the self-imposition of risks by a collective such as a nation or a transnational unit. On the other hand, it can be conceptualised as an act by which an authority imposes risks on others, e. g. on those parts of the population that are positively or negatively affected by the respective technology. In the first case, the decision to carry out the activity follows the decision theoretical model of subjective rationality. Since the deciding agent is identical with the agent who bears the benefits and the risks of the decision, the question is how to optimize the relation between benefits and risks given the preferences of the collective agent. In the second case, the appropriate model is the model of justified imposition of risks on others. The question is no longer a question of subjective rationality but a question of ethics. The question is whether it is morally legitimate for the authority to impose risks on others who may have preferences widely different from those deciding on or carrying out the risky activity.

If one adopts the latter, individualistic, point of view, as I think we should, it is plausible to take account of all relevant preferences of those affected by a technology, including their risk preferences. Even if the agent himself is a Bayesian, convinced that the best thing is to choose the option by which the expected value for all affected by the option is maximized in the long run, it is doubtful whether he is justified in generalizing this preference and to impose risk profiles on others which they, from their own risk preferences, want to steer clear of. As soon as others are affected by the agent’s choices the question arises whether it is legitimate to orient the imposition of risks exclusively on one’s own risk preference. Even if the agent, as far as he is concerned, is perfectly willing to have risks imposed on him in accordance with his own risk preferences, it is doubtful whether this legitimizes imposing corresponding risks on others. After all, it is not usually the case that we

are allowed to impose on others what we allow others to impose on us. (Think of G. B. Shaw's travesty of the Golden Rule: "Do not do unto others as you would be done by them. Their tastes might be different.") A physician who, as far he is concerned, would be willing to undergo a certain risky operation, cannot assume a priori that his patients share his risk preference and prefer the risky operation to a more conservative treatment involving less risks and less chances. If his obligation, in deciding about how to proceed, is to respect the preferences of his patients, it is part of the same obligation to respect their risk preferences.

It is a well-known fact that as far as the risk profile of technological options include risks of a certain severity, the risk attitudes prevailing in the general public are risk-averse rather than risk-neutral. This fact constitutes a further reason to modify the Bayesian approach in the direction of an approach that attaches more negative weight to adverse events and restricts the balancing of risks and benefits, without, however, unduly obstructing technological progress.

5 Conclusion

Risk is inherently a value concept and cannot be analyzed in purely descriptive terms. This is one reason why emotions in a broad sense including emotional attitudes are a central component in the assessment of the risks of technological options. Likewise, emotional factors play a part in the explicit or implicit valuation of the distribution of benefits and risks, in risk preference and in the characteristics shown by psychologists of risk to contribute to the acceptance or non-acceptance of risky technologies. Not all of these emotional factors are compatible with rationality. For reasons of economy and limitation of resources, emotions, just as perceptions, make use of simplifying heuristics that are often useful and sometimes misleading.

In this article, I have mainly dealt with emotional factors for which it is plausible to assume that they are compatible with rationality at least to a certain extent: voluntariness, control, potential for catastrophe, and uncertainty. These factors, however, are only a selection from the "qualitative characteristics" that have been found to determine the popular perception and acceptance of risks. Other factors in this list are, I think, not amenable to a reconstruction in the Bayesian or any other model of rationality. They are "emotional" not only in the sense that they are rooted in spontaneous and non-cognitive tendencies but also in the polemical sense designed to deny them intellectual respectability. Rather than useful heuristics in situations where the tools of formal analysis fail to be helpful, they mislead our thinking and misdirect our actions. Among these are: 1. symbolic values, 2. salience; 3. familiarity, and 4. naturalness. *Symbolic content* seems to play a significant role in the valuation of risk. For example, energy production from nuclear fission is inevitably associated with the nuclear bomb and solar energy with the life-giving role of the sun, so that it appears "natural" that the latter is less risky than the former. *Salience* is a factor in the assessment of technological risks as it is a factor in the individual's perception and evaluation of possible diseases. The frequencies of dramatic and sensational events are overrated, the frequencies of trivial events are underestimated (Slovic

1978, p. 100), with the consequence that people and politicians are more prepared to invest in security against murder and terrorism than in everyday causes of death like coronary infarction and infections from hospitalisation. *Familiarity*, again, makes that we hardly worry about risks that have become habitual features of our life-world even if they are far more substantial than less familiar risks. Interestingly, the emotional attitude underlying familiarity corresponds to the *absence* of emotion in the episodic sense. In a sense, we react to risks that have become familiar with less emotion than would be appropriate. In this way, deaths and injuries from car traffic have become familiar, whereas deaths and injuries by radiation have not. Without the factor of familiarity it seems difficult to explain that energy production by burning coal, with 15,000 deaths in German coal mines since 1948, is widely accepted whereas energy from nuclear power, with 0 deaths in German reactors, is not. It may be thought that more familiar risks are easier to tolerate because society has had time to adapt to these risks and to establish corresponding means to come to terms with them. This consideration, however, is hardly relevant because mechanisms and institutions to control and to correct these risks are already part of the calculation. The existence of a fire brigade is already part of the overall risks of fire, the institution of hospitals already part of the overall risks of car traffic. In consequence, the fact that more familiar risks are more easily accepted than unfamiliar ones has to be explained by habituation effects and cannot be rationalised along the lines of the dimension of certainty. At last, *naturalness* is an important factor in the acceptance of risks, for which, again, it is difficult to see how it can be interpreted as rationally defensible (cf. Hansson 2003). Natural causes of harm are given what might be called a “nature bonus”. Natural harms are less feared than anthropogenic harms, possibly because there is nobody in particular to blame for inflicting it. One cancer patient dying from the radioactivity emitted by a nuclear power plant will attract more attention than ten or hundred patients dying from natural radiation. Again, in the preference for natural above technical or other anthropogenic risks, emotions seem to play a central part, possibly due to evolutionary constraints such as the impossibility, over long periods in the history of mankind, to control natural risks (cf. Birnbacher 2006, p. 21 ff.).

All in all, then, emotions are a mixed blessing – in the assessment and acceptance of risks no less than in other domains of life.

References

- Birnbacher, D. 2003. Can discounting be justified? *International Journal of Sustainable Development* 6: 42–51.
- Birnbacher, D. 2006. *Natürlichkeit*. Berlin/New York: de Gruyter.
- Gigerenzer, G., and R. Selten. 2001. Rethinking rationality. In *Bounded Rationality. The Adaptive Toolbox*. G. Gigerenzer, and R. Selten, eds., 1–12, Cambridge, MA: The MIT Press.
- Hansson, S. O. 2003. Are natural risks less dangerous than technological risks? *Philosophia Naturalis* 40: 43–54.
- Harsanyi, J. C. 1978. Bayesian decision theory and utilitarian ethics. *Economics and Ethics* 68: 223–228.

- Johnson, E. J., and A. Tversky. 1983. Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology* 45: 20–31.
- Jungermann, H., and P. Slovic. 1993. Charakteristika individueller Risikowahrnehmung. In *Risikante Technologien, Reflexion und Regulation*. W. Krohn, and G. Krücken, eds., 79–100, Frankfurt/M: Suhrkamp.
- Kasperson, R. E., et al. 1988. The social amplification of risk. A conceptual framework. *Risk Analysis* 8: 177–187.
- Kates, R. W. 1981. *Risk Assessment of Environmental Hazards*. Chichester: John Wiley & Sons.
- Keeney, R. L. 1984. Ethics, decision analysis, and public policy. *Risk Analysis* 4: 117–129.
- Leist, A., and P. Schaber. 1995. Ethische Überlegungen zu Schaden, Risiko und Unsicherheit. In *Risikobewertung im Energiebereich*. M. Berg et al., ed., 47–70, Zürich: Verlag der Fachvereine.
- Nord, E. 1999. *Cost-Value Analysis in Health Care, Making Sense of QALYs*. New York: Oxford University Press.
- Posner, R. A. 2004. *Catastrophe, Risk and Response*. Cambridge: Cambridge University Press.
- Renn, O. 1991. Risk communication and the social amplification of risk. In *Communicating Risk to the Public*. R. E. Kasperson, and P. J. Stallen, eds., 287–324, The Netherlands: Dordrecht.
- Renn, O. 2008. Concepts of risk, an interdisciplinary review. *Gaia* 17(50–66): 196–204.
- Renn, O., and M. Zwick. 1997. *Risiko- und Technikakzeptanz*. Berlin: Springer.
- Rescher, N. 1983. *Risk. A Philosophical Introduction to the Theory of Risk Evaluation and Management*. Lanham, MD: University Press of America.
- Slovic, P. 1978. Judgement, choice and societal risk taking. In *Judgement and Decision in Public Policy Formation*. K. A. Hammond, ed., 98–111, Boulder, CO: Westview Press.
- Slovic, P., B. Fischhoff, and S. Lichtenstein. 1979. Rating the risks. *Environment* 21: 14–39.
- Starr, C. 1969. Social benefit versus technological risk. *Science* 165: 1232–1238.
- Starr, C., R. Rudman, and C. Whipple. 1976. Philosophical basis for risk analysis. *Annual Review of Energy* 21: 629–662.
- Trapp, R. W. 1988. “Nicht-klassischer” Utilitarismus. *Eine Theorie der Gerechtigkeit*. Frankfurt/M: Klostermann.
- Tversky, A. 1975. A critique of expected utility theory, descriptive and normative considerations. *Erkenntnis* 9: 163–173.
- Tversky, A., and D. Kahneman. 1974. Judgement under uncertainty, heuristics and biases. *Science* 185: 1124–1131.
- Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211: 453–458.

Emotions Involved in Risk Perception: From Sociological and Psychological Risk Studies Towards a Neosentimentalist Meta-Ethics

Felicitas Kraemer

1 The Sociology of Risk: Risk Objectivists Versus Risk Constructivists

Since the 1970s, there have been ongoing discussions about the nature of risk, triggered by the analysis of societal attitudes towards new technologies, for instance of nuclear power, nuclear weapons, technologies causing environmental damage and new developments in biotechnologies. There were two competing definitions of risk, supported by the risk “objectivists” and the risk “constructivists”.

On the objectivist side, there was the traditional statistic definition of risk as an objective entity, i.e. the probability of an event multiplied by the estimated severity of consequences. It served as the basis for insurance agencies to calculate their gains and losses. (Cf. Manes 1913, pp. 21 ff.; Schöpfer 1976; Peters 1959; all cited in Krohn and Krücken 1993, pp. 16 ff. Cf. for a hybrid account that also included objectivist elements Beck 1986, 1988). On the other hand, there were the so-called risk constructivists. They primarily observed a growing societal sensitivity to risk. They regarded risk as mainly constituted by and dependent on societal perception. (Krohn and Krücken 1993). In Sections 1 and 2, I will give a brief overview of the sociology of risk, focusing on the role of emotions. In Section 3, I will critically discuss ideas by Sjöberg. In Section 4, I will propose a moderate philosophical, meta-ethical form of constructivism with respect to risk that can be called “neosentimentalism”. In Section 5 I will discuss the practical implications of my view.

An interesting approach within the constructivist tradition stems from Niklas Luhmann who pointed out that the antonym to “risk” is not safety, but “danger.” For him, danger is a state the individual is in that someone else has caused, and something the individual is involuntarily exposed to. Risk, however, is to be understood as something an individual actively takes, or intentionally imposes

F. Kraemer (✉)

Faculty of Innovation Management and Industrial Design, Philosophy & Ethics, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: f.Kraemer@tue.nl

on someone else (Luhmann 1993). The strong socio-constructivist developments in the 1970s had already undermined the monopoly of the risk objectivists to a certain extent. Additionally, there was an intrinsic problem that weakened the objectivist position even more and made it seem inadequate: New technologies brought about unprecedented uncertainty and scientific unpredictability. Think for example of nuclear accidents: neither their probability nor their harmful consequences are properly assessable on an objective scale. There is not enough expert knowledge available, and the only way to find out more about these accidents would actually be to intentionally cause them (Taubert and Kraemer 2007). In this respect, it is interesting to note that even the Chernobyl accident went back to a willfully conducted experiment to test the cooling circuit of the reactor to assess risks to prevent emergencies.

As Dieter Birnbacher has pointed out, some risks involved in biotechnologies also raise difficulties for “objective” risk assessment (Birnbacher 1993). For instance, the genetic manipulation of organisms is a high risk enterprise for which “objective” data are not sufficiently available. Especially for such new technologies, there seems to be no uncontroversial “objective” point of view to properly assess them.

Hence, on closer inspection, it turns out that the objectivist standard fails and that therefore, constructivism seems to be the most straightforward alternative. Recently, especially research done on the Chernobyl case shows that there is no objectively accountable point of view on the magnitude of damage and of the side-effects brought about by high risk enterprises (IPPNW 2007, p. 10, SSK (Strahlenschutzkommission) 2006; Lengfelder 1990, cf; also Taubert and Kraemer 2007). This is mirrored by the fact that for instance nuclear plants cannot contract insurance for the consequences an accident brings about. There is certainly no insurance company in the world that, apart from simple contents insurance, would be willing to cover the environmental damage caused by a nuclear accident. There is no objective account of the magnitude of the possible damage available, or it is so gigantic that nobody could pay for it (cf. Taubert and Kraemer 2007).

In the vein of opposing a simplistic objectivism of risk, already in 1982, the cultural anthropologists Mary Douglas and Aaron Wildavsky published their influential socio-constructivist analysis of risk (Douglas and Wildavsky 1982). They raised the question of whether the number of dangers in the age of technology actually increases, or whether it is just the fear of risks associated with new technologies that grows. There is no simple answer available. On the one hand, there are many new sources of risk. On the other hand, our life has become much safer than before, and there is an increased expectation of life thanks to new technologies. The US statistics Wildavsky and Douglas referred to 25 years ago showed that most people were afraid of wars, crime, environmental disasters, and loss of wealth. It turns out that an individual cannot evaluate the full range of risks that it is exposed to. Therefore individuals search for institutions to select and prioritize the most important risks for them. Accordingly, emotions towards risk turn out to be social constructs. Individuals cognitively learn from a risk culture or group what to be most

afraid of, and only then, based on this shared knowledge, they actually develop the respective emotion, for example fear. Douglas' and Wildavsky's most important result was that even the experts and risk managers in a society are members of risk cultures. Therefore, like lay people, they selectively pick out special kinds of risk they are aware of – and not others. They do not have a privileged objective perspective on risks. There is no objective “God's eye view” on risk. In the light of this, a naive objectivist approach towards risk seems to stand on shaky ground (Douglas and Wildavski 1982).

2 Qualitative Risk Perception

These risk studies in the 1970s were mainly conducted by sociologists and psychologists. They indicate that a social constructivist notion of risk was on the rise. There are two factors that are relevant for this approach: First, it had turned out that the objectivist approach, understanding risk as an objective entity, did not pay justice to new technologies that lacked objectifiable risk data. Second, it did not do justice to the growing participatory consciousness of the public that wanted to have a voice in risk assessment and that brought their emotions of fear and uneasiness into play, demanding a policy that is sensitive to their concerns and feelings.

This constructivist turn brought about numerous new studies in the sociology and psychology of risk perception, putting emphasis on subjective, qualitative factors. Among them were:

- The catastrophic potential of an event: A risk is perceived as higher if it has extremely severe, irreversible, and wide-ranging consequences.
- Personal affectedness: a person regards a situation that affects herself as much more risky than if others are concerned.
- Perceived controllability, that is for instance the idea that I am an extraordinary good driver and have full control of the situation, so nothing will happen to me; the risk is perceived as lower.
- Involuntariness of exposure: A risk is perceived as higher if people undergo it involuntarily.
- The naturalness of the sources of risk makes a risk seem lower, in contrast to cases where an agent is responsible for a risk by intentionally interfering with natural processes like in e.g. anthropogenic climate change.

This list is still incomplete. Paul Slovic, for instance, has a list of 18 features that are commonly attributed to risk assessment by lay people (Slovic 2000, cf.; Roeser 2007, p. 3). In 1969, Charles Starr created the so-called psychometric model of risk perception. It highlighted the level of risk tolerance explained by the dread and novelty of technologies. For instance, nuclear power is ranked highly on the scales of dread and novelty. Therefore many people were opposed to this technology (Starr 1969).

Some recent research by Gaskell et al. (2003 and Siegrist (2003, both cited in Townsend (2006, p. 130) that Ellen Townsend discusses in her paper “Risk Perception and GM Food”, show that the assessment of risk depends to a great extent on the emotional associations people connect with certain stigmatized mental images that arise from thoughts of “genetically modified food” (cf. Townsend 2006, p. 130 f.; Gaskell et al. 2003; Siegrist 2003; cited in Townsend 2006, p. 130). If one applies to this a term from ethics, one could describe this as the “yuck-factor”, i.e. the gut reactions of repugnance against genetically modified foods. According to Gaskell et al. (2003 these gut reactions raise mental images of “infection” and “monstrosity” that are commonly associated with genetically modified organisms (Gaskell et al. 2003; cited in Townsend 2006, p. 130). There seems also to be a link between *moral assessment* and risk assessment (Gaskell et al. 1997; cited in Townsend 2006, p. 130). Risk attitudes are based on a collection of anxieties about unforeseen dangers that may be involved in a range of technologies that are commonly perceived to be “unnatural” (Gaskell et al. 1997; cited in Townsend, p. 130.) Accordingly, an enterprise or technology was regarded as most risky when it was regarded as “unethical” (cf. Townsend 2006, p. 130, in referring to a study by Ferguson et al. 2001). As Townsend reports about results of of Ferguson et al., “in the only study that has specifically examined the feelings of dread with GM food”, the result was that “of the 20 concerns investigated in the study including, e.g. human cloning, CJD, biological warfare and car crashes, GM food was the least dreaded” (Townsend 2006, p. 130, referring to Ferguson et al. 2001).

In a similar vein, the social psychologists Wiedemann and Schütz show an interconnection between moral assessment and risk estimation. In their empirical studies they investigate the pre-eminent role of so-called risk-stories that are regularly created around technological accidents (Wiedemann and Schütz 2000; cited in Wiedemann and Brüggemann 2000, pp. 13–14). People assess the potential harmful consequences of certain technologies intuitively as more dangerous if the story told elicits feelings of empathy with the agent, e.g. if he just had bad moral luck. In contrast, they judge the same risk as more severe and dangerous if they feel rage against the morally inappropriate behaviour of the responsible agent, e.g. if the accident was caused intentionally. Additionally, the subjects were more forgiving with respect to small firms than to big businesses, and their risk assessment exposed a higher estimation of risk for the latter (Wiedemann and Schütz 2000; cited in Wiedemann and Brüggemann 2000, pp. 13–14). Wiedemann’s and Schütz’ results can be interpreted as hinting towards the fact that the negative moral emotion of indignation is strongly correlated with the negative epistemic emotion of fear. The positive moral emotion of empathy, in turn, seems likely to elicit the positive epistemic emotion of trust that goes along with a lowered perception of fear.

In summary, numerous studies have pointed out that emotions like fear and trust are epistemic factors that are inherent in qualitative risk perception, and that these emotions oftentimes have a normative quality insofar as they are linked to values (judging that something is good or bad). Nevertheless, there seems to be a new objectivism or anti-sentimentalism on the rise with regard to risk.

3 A New “Antisentimentalism” in Recent Risk Studies?

In March 2006, a leading journal, the *Journal of Risk Research*, featured a special issue on “Risk and Affect.” The editorial article was authored by the Swedish social psychologist Lennart Sjöberg. I will first give an overview of his argument and then critically discuss his ideas.

In his paper entitled “Will the Real Meaning of Affect Please Stand Up?”, Sjöberg states that in previous risk research, it was “widely believed that affect plays an important role in risk perception, and that such perception is mainly governed by emotional processes.” (Sjöberg 2006, p. 101). In contrast, Sjöberg maintains that this “belief is based on weak evidence, if the words affect and emotion are interpreted according to their dominating meanings in natural language, and to common usage in psychology at large” (Sjöberg 2006, p. 101).

In a *first* step, Sjöberg criticizes sentimentalism in risk studies by providing an analysis of the terms “affect” and “emotion” in psychology and everyday language. His conclusion is that the “word affect should be used to denote emotion” (Sjöberg 2006, p. 101).

In a *second* step, Sjöberg points out that there is weak *empirical* evidence for the hypothesis that risk perception is based on emotional processes. In this vein, he mainly opposes Fischhoff’s and Slovic’s assumption of the pre-eminence of the emotion-based so-called “Dread Factor”. (Fischhoff et al. 1978; Slovic 1987; cited in Sjöberg 2006, pp. 105 f.) To support this assumption, Sjöberg refers to the findings by Graham (2001; cited in Slovic, p. 105 and cited in Sjöberg 2006, p. 105). Graham shows that Dread is not a homogeneous concept, and that it has no clear relation to emotions. Rather, is a heterogeneous notion that includes elements such as severity of consequences and other elements that have nothing to do with emotions (Graham 2001; cited in Sjöberg 2006, p. 105). The other elements Graham mentions are fatality, globality, involuntariness, uncontrollability, unfairness, catastrophic versus unclustered victims, impact on future generations, increase, irreducibility. In Sjöberg’s eyes, all these features lack a clear relation to emotional experience (*ibid.*). For Sjöberg, it is not Dread, but the rational calculation of the *Severity of Consequences* that are major factors in risk assessment. It is only in the end that emotions arise from a rational consideration of severe consequences. Emotions are mere epiphenomena of an otherwise rational process of deliberation. One of Sjöberg’s examples in an earlier paper is nuclear waste as a hazard for future generations. The concern about this fact, for Sjöberg, is not an “emotional” but a highly rational response. It more or less mirrors the “objective reality” of a risk scenario (Sjöberg 2003). Sjöberg adds that there is an established connection between “liking” and “risk”, whereas in his eyes, “liking” is not an emotional state. From Sjöberg’s editorial article, it does not become entirely clear what “liking” really is. (Sjöberg 2006, pp. 103, 106 f.; Sjöberg refers to Finuncane et al. 2000). Anyways, for him, the connection of risk to liking is no clear evidence for a connection between emotions and risk.

In a *third* step, Sjöberg explains the strategic reason why he opposes the picture of an emotionally driven public. His worry is that it could have unwanted

policy implications. Already in a 2003 paper, Sjöberg had criticized the literature on risk research since the 1980s since it portrayed people's policy attitudes mainly as emotionally motivated (Sjöberg 2003).

In his eyes, a sentimentalism of risk, as I would call it, implies that one should lower the level of public participation when it comes to judgement about risk exposition and should rather rely on experts. Those who picture lay persons as highly emotional run the danger of presenting them as unreliable and biased when it comes to risk assessment. Sjöberg wants the public to look more rational because he wants to enhance public participation and to avoid expertocracy (Sjöberg 2006, p. 101). He thinks that empirical research shows that emotional factors play only a minor role in risk perception.

In simplified terms, I will call Sjöberg's approach "anti-sentimentalist" with regard to risk. In what follows, I will discuss Sjöberg's approach from three aspects along his three steps:

First, from a philosophical perspective, it is difficult to understand Sjöberg's criticism of the use of the terms "affect" and "emotion" and his demand to use them as synonyms. In contrast, it seems that Sjöberg identifies emotions with simple body-centered, primitive "gut reactions" that for instance follow upon the experience of dread: "The term dread clearly suggests that people have a 'gut reaction' to a hazard and that such a reaction is the main part of the dynamics of their concern" (Sjöberg 2006, p. 101).

However, many emotion theorists, among them cognitivists such as Solomon (2003) and Nussbaum (2001), would reject such a reduction of emotions to physiological gut reactions. It would go beyond the scope of this paper to discuss all potential philosophical and psychological emotion theories. A helpful overview can be found in De Sousa (2003/2007) who discusses feeling theories, cognitivist theories, perceptual theories, psychological and evolutionary approaches of emotions (De Sousa 2003/2007). Emotions can be understood as complex mental states that have a qualitative, affective side as well as an intentional, cognitive structure. According to Peter Goldie, emotions are "feelings towards". They have an intentional structure, i.e. a content or an object that is qualitatively perceived (Goldie 2009, p. 115). This shows that, from a philosophical perspective, Sjöberg's account of emotions as mere gut reactions would not be shared by many authors. Another way of avoiding his identification of emotions with gut reactions are perceptual or sensibility theories of emotions. For the purpose of this contribution, I will advocate a sensibility theory of emotions that regards values as secondary properties which are actualized by certain emotions (cf. Prinz 2007, p. 108, Wiggins 1998; McDowell 1997). Such a view was recently defended by David Wiggins and John McDowell. I will elaborate on Wiggins' approach in Sections 4 and 5.

At this point, my two comments on Sjöberg's first step are: He works with an oversimplified picture of what an emotion is when he understands it as a bodily reaction. This does not do justice to the many sophisticated accounts of what emotions are that can be found in the literature and makes emotions much more "irrational" than they really are. Further, even in the face of a large variety of emotion theories, there is a minimal consensus between numerous emotion theorists: Emotions

have a phenomenal quality, they are felt in a certain way, and are closely related to value perception (good and bad, which is oftentimes called “appraisal”; cf. the appraisal theories by cf. Scherer 1999). From the sociological and psychological studies quoted in Sections 1 and 2, it rather seems that this qualitative experience does in deed play a key role in risk assessment. Take a person who utters the sentence: “I am afraid of the health hazards that the storage of nuclear waste could cause to my children and grandchildren and therefore oppose nuclear energy/the insecure storage of nuclear waste/etc.” It makes sense that severe consequences are qualitatively perceived in the sense of negative appraisal.

Second, and interrelated with this point, Sjöberg disregards the intrinsic interconnectedness of emotions on the one hand and values on the other. As he puts it:

People react emotionally, yes, but their policy attitudes are dependent on a host of factors which may more correctly be named ideological and value loaded rather than emotional. Values and attitudes are one thing, emotions another. It is common in heated debates to accuse the opposite party of being “emotional”, hence beyond rational appeals and driven by strong forces which have nothing to do with a rational approach to policy. This superficial and rhetorical stance should not be embraced by risk researchers. What we find is that people have different beliefs and values, and that all variations have relatively little to do with emotions. (Sjöberg 2003, p. 108).

Sjöberg does not give any argument for his splitting up of emotions and values. Further, from a philosophical perspective, it remains somewhat unclear what he means by “values” here. However, in a previous paper co-authored with Elisabeth Engelberg, Sjöberg adapts the definition of values provided by the social psychologist Shalom Schwartz. Schwartz regarded values as “the criteria people use to select and justify actions and to evaluate people (including the self) and events” (Schwartz 1992, p. 1; cited in Sjöberg and Engelberg 2005, p. 327). Sjöberg and Engelberg point out that in their understanding, values are “judgments similar to the one used in the measurement of attitudes, of a general or abstract concept”, and that examples of such concepts “are freedom and equality” (Sjöberg and Engelberg 2005, p. 330).

In a paper co-authored with Britt Drotts-Sjöberg, Sjöberg states that values are ideas that are appreciated by persons and give them orientation in their lives (cf. Sjöberg and Drotts-Sjöberg 1997). Accordingly, among these “value dimensions” are “a main dimension of individualism (personal success) vs. collectivism (solidarity). Further, repository opponents often relied on values or explanations emphasizing tradition, small-scale establishments, personal control, the need for high level security, risk for future generations, and the importance of preserving nature and keeping the wilderness intact” (cf. Sjöberg and Drotts-Sjöberg 1997, pp. 115 f.).

To sum up this discussion, Sjöberg regards values as rational judgments and as objects that are worth being cherished. This enables him to distinguish between values as judgments on the one hand and emotions on the other, the latter being merely gut reactions for him. Therefore, his terminology differs sharply from the one that is widely used in philosophical discussions about emotional cognitivism. So-called cognitivist theories of emotions have a very different understanding of emotions. For instance the cognitivists Martha Nussbaum and Robert C. Solomon regard emotions

as judgments (Solomon 2003; Nussbaum 2001) and would therefore certainly reject Sjöberg's idea of a gap between emotions and judgments.

In contrast to Sjöberg, according to various philosophers, there are strong ties between emotions and values, and they are at least twofold. On the one hand, following realist authors such as Max Scheler in emotion theory, (Scheler 1980) it could be possible that we perceive values via emotions. This comes close to the emotion theory Sabine Roeser supports in the context of risk perception (Roeser 2006, 2009, 2010). On the other hand, there is the sentimentalist tradition following authors such as David Hume. In this understanding, emotions generate values. For beings that are unable to experience emotions, the world would not provide any values. I will support such a sentimentalist or rather neosentimentalist position in Chapter 4 of this paper. Sjöberg's approach, however, that splits up emotions from values and declares them two different categories, does not fit in either of these theories. His ideas fit with rationalist views in metaethics (such as defended by Kantians), but from the point of view of philosophy of emotions, his ideas seems implausible. Philosophers of emotions emphasize that without emotions we could not appreciate values. In the context of risk, these are values such as egalitarianism and the preservation of nature.

Third, Sjöberg regards "emotions" as irrational forces. Therefore, he chides risk-sentimentalists for their making the public look emotional and thus "irrational". From a philosophical perspective, this is implausible as well. Following authors such as Antonio Damasio and Ronald De Sousa, it becomes clear that emotions and rationality are closely intertwined (De Sousa 1990; see also Roeser 2010; Chapter "Emotional Reflection about Risks" by Roeser, in this volume). In contrast to authors such as Sjöberg, the sociologist Charles Perrow pointed out the so-called "social rationality" of lay people. Perrow turns the tables round. For him, the alleged rationality of experts is irrational insofar as it is blind to non-quantifiable harm like for instance the loss of trust in institutions. The emotionality of lay people, however, can be seen as rational insofar as it intuitively captures the non-quantifiable aspects of risky technology. In my eyes, Perrow's position does justice to the rational content and evaluative character of emotions whereas Sjöberg misses it (Perrow 1984).

To sum up this discussion, Sjöberg's analysis ignores the importance of emotions for qualitative risk perception. In my eyes, both is possible: To allow emotions to come into the picture of risk perception, and at the same time support the participation of the public to policy making. Emotions are needed in order to see the importance of qualitative risk factors. Based on a different view of emotions, the contradiction that Sjöberg sees can easily be avoided.

4 Some Metaethical Implications

In the first part of this paper, I contrasted an objectivist perspective on risk with a socio-constructivist one. The abovementioned socio-constructivist authors from psychology and sociology shared the assumption that risk must be more than an

objective fact. A risk is something that is *perceived* by a subject, a group or a society *as dangerous or threatening*. Here, emotions come into play as a way of perceiving something as risky, for instance if something is experienced as a dread.

If we look at these two approaches (the objectivist and the constructivist) from a philosophical point of view, it turns out that they mirror two different positions in metaethics that I will present in simplified terms. On the one hand, there is a *realism* of properties. According to realism, properties such as “dangerousness” or “beautiful” really exist in the world or in a transcendent realm. A sub-category of these properties are moral values such as “good” or “bad”. For the realists, moral values exist independently of a subject that perceives them. Accordingly, the property of being risky really exists out there in the world, as a property of certain objects or events. Most realists think that we understand objective values through reason. On the other hand, there are constructivists or subjectivists of the Humean kind. For them, properties exist depending on the subject that perceives them. This position is oftentimes referred to as “subjectivism” or “projectivism.” For instance, in his *Treatise*, Hume claims that the beauty of a pillar lies in the eye of the beholder, and the vice inherent in a murder is projected into the scene by the person who watches it (Hume 1978, p. 65, cf.; Hume 1975, esp. pp. 110–120). According to Hume, emotions play a crucial part for values. They engender or bring about values. The adoring and joyful emotions of the beholder make the pillar beautiful. Without an emotional spectator, there would be no such thing as beauty in the pillar. Similarly, one could state that without the perceptive emotions of certain individuals, there would be nothing in the world that could be understood as risky. Taken this way, the predicate “risky” can be understood as a response-dependent property. Its existence depends on its being qualitatively perceived. Descriptively speaking, here is where emotions come into play as constructive, epistemological factors. Accordingly, following a Humean line of thought applied to risk, one could call this a “sentimentalist” perspective on risk.

As Sabine Roeser points out, there seems to be a third way of understanding the role of emotions in risk perception. She supports an intuitionism of values in which intuitions understood as emotions are capable of perceiving objective values. Thus, she combines the realistic background assumption that there are objective values with the idea that emotions play a crucial role in risk perception, since in this view, it is via emotions that we access objective properties, i.e. in this case *objective* moral values (Roeser 2006, 2010; also cf. Scheler 1980, esp. Chapter 2 V).

I agree that Sabine Roeser’s approach is a veritable alternative to the abovementioned two positions of the realists and constructivists or subjectivists. However, I do not share her intuitionism. It would go beyond the scope of this paper to explain in details all reasons for this. Here, I will only hint to the fact that the discussion of a realism of values has a long and controversial history. Among the reasons that lead many authors to a rejection of the realistic view is the fact that it relies on strong metaphysical assumptions about the nature and ontology of values (Mackie 1991, pp. 38 ff.).

In order to avoid commitments to a realistic ontology of values, I think it would be worth developing a sentimentalist account with respect of risk, or even a so-called

neosentimentalist. Since the 90s, there is a discussion going on over a renaissance of sentimentalism in metaethics that runs under the label of so-called “neosentimentalism”. In this debate about the interrelatedness of emotions and values, authors such as David Wiggins, John McDowell, Allan Gibbard and Simon Blackburn, to name only a few, play a leading role (D’Arms and Jacobson 2000a, pp. 722–748, cf., 2000b, pp. 65 ff., Gibbard 1990; Blackburn 2000; Wiggins 1998; cf. Steinfath 2001). Although they come from different metaethical backgrounds, they all support versions of the sentimentalist idea of a response-dependence of values and therefore have created a revival of traditional sentimentalist ideas in the form of neosentimentalism (D’Arms and Jacobson, 2000a). Neosentimentalists focus on the question of the *appropriateness* of emotions and thereby try to avoid relativistic and subjectivistic pitfalls of the classical Humean sentimentalism (Nichols 2004, pp. 65–70). In the Humean picture, there is no remedy for a subjectivism and relativism of values that necessarily comes along with his sentimentalist approach. In contrast, the neosentimentalists basically follow Adam Smith’s idea of the “impartial spectator”. The idea of such an impartial spectator goes back to a thought experiment about an ideal person who has appropriate emotions (Smith 1759, pp. 110, 113, 167, cf. Gibbard 2005, p. 277).

Due to space restrictions, I cannot go into detail about a possible neosentimentalist metaethics of risk that could be developed in the future. However, a neosentimentalist account of risk seems to hint into the right direction. It accommodates the socio-constructivist idea risks studies support since the 1970s, transformed to a metaethical level. As pointed out in the first part of this paper, opposing a merely objective, quantifiable idea of risk, for the constructivists, risk is something that is perceived as such by a subject or a group of subjects. Similarly, for a neosentimentalist understanding of risk, to be regarded as risky, a certain object or event has to be emotionally perceived as risky by an individual or group.

It is important to note that one has to distinguish between the cognitivist group of neosentimentalists exemplified by Wiggins and McDowell on the one hand and the non-cognitivist group embodied by Blackburn and Gibbard on the other (cf. Prinz 2007, p. 108). Only the latter would possibly embrace the idea of a socio-constructivism that is mirrored in their metaethics. We cannot go into detail about these exegetic questions here. However, the main difference between such a neosentimentalist notion of risk and Sabine Roeser’s realist account of risk is that Roeser argues for the thesis that emotions can *inform* us about objective moral salience (cf. Roeser 2009, 2007, p. 10). In contrast, I support the thesis that *emotions constitute salience*. In the neosentimentalist framework, risk is a response-dependent notion. Emotions do not track values and properties as they do in the realist picture, but constitute them. In contrast to Humean sentimentalists, for neosentimentalists, however, only those emotions constitute values and properties that are appropriate, i.e. that are in line with the emotions an ideal impartial spectator would have. According to a realist point of view, an ideal spectator would track objective moral truths in a reliable way, establishing a relation of *correspondence* between emotions and objective values. From a constructivist point of view, however, the standard by which emotions are determined as ideal and appropriate consists in the criterion of

coherence of different emotions with each other. Here, I cannot embark upon an in-depth interpretation of Adam Smith's ideas. Rather, in what follows, I will briefly discuss the idea of emotional appropriateness understood as coherence with reference to the discussion about *neosentimentalism*. This discussion largely refers back to Adam Smith (Gibbard 2005).

I can only briefly hint at the view of two neosentimentalist authors. First, Allan Gibbard in his book *Wise Choices, Apt Feelings* argues for a naturalized account of appropriateness. For him, the emotions of an individual are appropriate if they are correlated successfully with the emotions of others and form a coherent web of emotions. In the end, emotions are appropriate if they enable effective social cooperation and are thus useful in terms of reproductive success (Gibbard 2005, pp. 278–290).

A second argument is offered by David Wiggins in his essay "A Sensible Subjectivism" (Wiggins 1998). His strategy seems more promising to me, because it provides a non-naturalistic explanation for the appropriateness of emotions. It has two elements. First, Wiggins supports a *sensibility*-theory of values. Values are secondary properties which are response-dependent, i.e. they are actualized only via emotions, but are nevertheless based on certain dispositional structures of the object. The object "deserves" certain emotions and not others, as David Wiggins puts it (Wiggins 1998, p. 210). This means that the appropriateness of emotions is defined with regard to dispositional properties of the object itself. The second aspect of Wiggins' theory, like Gibbard's, is based on coherence. Wiggins tells a story about the socio-historical development of values in which values are constituted via emotions that have become socially accepted and that are inter-subjectively shared. The coherence of this inter-subjectively stabilized net of emotions establishes a second criterion for the appropriateness of emotions. Over time, such a web of established emotional responses constitutes the structure of the object's dispositional properties. The dispositional properties match more and more with the socially accepted coherent web of emotions and become enriched by new emotions. Vice versa, in the course of human civilization, the qualitatively enriched objects evoke new emotional responses. In contrast to a realism of values, Wiggins talks about a "non-vicious circle" of co-creation of values and emotions (Wiggins 1998, p. 212).

If one accepts Wiggins' neosentimentalist framework and applies it to emotions involved in risk perception, similar to secondary properties, these emotions can be regarded as co-constituting and "enriching" the object over time in its qualities. Further they must accord with the criterion of coherence in order to count as appropriate. For instance, in order to count as appropriate, the emotion of fear an individual or group experiences in the face of a certain new technology has to be shared by a growing number of people. The emotion has to become a socially established one to count as appropriate and to be taken into consideration when it comes to normative assessment about the acceptability of risks.

Even in the face of these metaethical differences between Roeser's intuitionism with respect to risk and my suggested neosentimentalism, the normative, practical results and implications for policy-makers seem to be more or less the same. Sabine Roeser argues that more attention should be given to our emotional reactions

with respect to risk in order to complement the scientific, quantitative notion of risk (Roeser 2007). Analogously, I will argue in the following section that emotions have to be taken into account when it comes to the normative assessment of risk within the framework of a recipient-oriented approach.

5 Normative Questions: Risk and Emotional Damage

So far, when considering metaethics, we have dealt with the epistemic role of emotions in risk assessment. But how are things with reference to normative ethics? What could a (neo-)sentimentalist account of risk contribute to a normative ethics of risk? There are several central questions of normative risk ethics: First, what kind of risk exposure of others is morally legitimate, oftentimes phrased as the question “How safe is safe enough”? (Birnbacher 1999, p. 137). Second, does emotional damage add to the total amount of harmful consequences? With respect to the first questions, there is a plethora of literature by authors such as Nicholas Rescher (1983), Kristin Shrader-Frechette (1991), Hansson (2004, Birnbacher (1993, 1999), and Sabine Roeser (2006, 2007, 2009, 2010), to name only a few. In what follows, I will therefore focus on the latter question.

In the previous chapters, we considered the epistemic role of emotions in risk assessment. In contrast to an objectivist concept of risk, I supported a neosentimentalist account according to which risks are not merely “out there”, but to a certain extent are dependent on and constituted via emotional perception. This means that it makes sense to talk about risks only with respect to sentient beings that emotionally experience them as such. In this understanding, the quality of an object or event to be risky is a response-dependent property.

If one agrees with the assumption that risk is response-dependent with respect to emotions of fear, it is a natural step to say that one has to pay attention to emotions such as fear which constituted the risk when the question arises how safe is safe enough. One could object that it would be a genetic fallacy to state that from the epistemic relation (emotions constitute risk), a normative evaluation follows (we have to pay attention to emotions when it comes to risk assessment). However, even if one is well aware of this fallacy, there seems to at least be a correlation between the epistemic and the normative level: If we want to assess the *full level of risk a person is exposed to*, we have to include in our consideration the level of risk she emotionally experiences. In short: If we aim at an understanding of the degree of acceptable risk, we have to consider the level of her negative emotional involvement in the situation. For a person who has strong fears with respect to a certain technology, the amount of risk she finds acceptable will be much lower than for a person free from fear. Therefore, there seems to be a correlation between the epistemic and the normative level of consideration here. One could say that this follows already from the metaethical account in section 4. However, it is certainly wise to make sure that one does not commit a genetic fallacy when inferring normative conclusions from constitutional relations.

The idea that we have to include the feelings of those who are exposed to risks has found its expression in so-called recipient-oriented approaches in risk ethics. A procedural “recipient oriented” approach in risk ethics goes back to Otway and Pahner in the social sciences and was recently renewed by Dieter Birnbacher (Otway and Pahner 1979, Birnbacher 1999, pp. 139 ff.). It sheds some special light on the role of emotions in the assessment of risk *consequences*. According to Dieter Birnbacher, in the framework of this ethics, the agent has to pay special attention to emotional consequences of his risky behaviour (Birnbacher 1996). Therefore, the question of how much risk may be imposed on another person implies the consideration of *psychological, emotional damage*. Reaching back to Otway and Pahner (1979), Birnbacher states that fears, as irrational as they may be, should be included into a cost-benefit-analysis as negative posts. The decider has no right to impose his or her standard of rationality and risk attitude on others. In this vein, the fear of the victims morally adds to the damage done to them through other consequences. Like other uneasy emotions, fear severely lowers the quality of life of a person. An example already mentioned before are the feelings of repugnance against genetically modified foods (cf. Townsend 2006, pp. 130 f.; Gaskell et al. 2003; Siegrist 2003; cited in Townsend 2006, p. 130). According to the recipient oriented approach, it is morally illegitimate to inflict emotional pains on risk averse consumers. It lowers a person’s quality of life to a considerable extent if she is afraid of supposedly poisonous food for herself and her children. Speaking on the level of normative ethics, the respective policies should therefore be subject to ethical reassessment. Another case might be the Chernobyl accident. Among the emotional consequences of it was the severe loss of trust in regulatory authorities and in scientific experts who were proven unable to deal adequately with the arising problems. This emotional damage, the so-called “confidence crisis” in experts and democratic institutions, adds to the directly caused number of casualties and health hazards as for instance Stig Nohrstedt described for the case of Sweden (Nohrstedt 1991).

However, a question arising here is: Should we respect fear in each and every case? Why not distinguish between reasonable and unreasonable fear? A severe problem with the recipient-oriented position is that emotions involved in risk perception are sometimes utterly irrational. As empirical studies have shown, they sometimes distort and falsify “objective” risks. In the beginning I listed some qualitative features of risk perception. Emotions sometimes make an “objective” danger seem more dangerous than it actually is, and sometimes make it seem “unreasonably” lower than it actually is (cf. Roeser 2010). Therefore, emotions seem oftentimes chided as “biased” and unreasonable.

In first response to this, one could say that in many cases, it might be appropriate to even consider irrational emotions. Although, epistemically speaking, a certain person is wrong, a strictly recipient oriented approach would nevertheless demand that we pay respect to her feelings – as irrational as they may be.

For example, from our cultural perspective, we might consider it irrational that American Indians refuse to be photographed because they are afraid of damage done to their souls. We may regard their belief as mere superstition and as

epistemically inadequate. Nevertheless, we deem it morally appropriate to respect their repugnance towards this technology.

Such a response, however, seems to be valid only in those cases in which the costs of the avoidance of emotional harm do not exceed its emotional benefits. The photographer who has to abstain from taking pictures, in this case to prevent emotional damage, will certainly do this at no high cost, whereas the benefit of the American Indians exceeds the costs by far. In this simplistic example, the recipient oriented approach is justified.

However, the case will be different if there are high costs relating to the attention to emotional factors. As Dieter Birnbacher reports, in the technological literature, you often find complaints about allegedly “irrational” prevention devices to calm down the affective concerns of people. Examples are especially expensive safety technologies in the context of nuclear plants (cf. Birnbacher 1999). As Birnbacher and Otway and Pahner point out, fears, loss of trust and existential uncertainty are as real as other material damages, although they are harder to register and to quantify. Accordingly, they have to be taken as equally serious damages as injuries and diseases (cf. Otway and Pahner 1979; Birnbacher 1999). If the safety devices in question really help to reduce people’s fears, for a recipient oriented approach, their implementation is worth considering despite the high costs, because they help reducing emotional harm.

6 Outlook and Conclusion

Nevertheless, this latter example hints at a deeper problem: What we need here is a distinction between appropriate and inappropriate emotions that might enable us to make well-grounded decisions on a normative level. Which kind of fear, for instance, should be taken into account and which not when it comes to the assessment of damage? Future research has to show to what extent the *neosentimentalist* debate about the appropriateness of emotions could help here. A tentative answer might consist in a two level model of normative risk assessment. If we accept a recipient oriented approach, we could, to begin with, consider *all* emotions as contributing to the amount of damage, no matter whether we deem them appropriate or not. On a second level, however, we should have a closer look at those emotions that are controversial, e.g. at morally doubtful emotions such as racist fears, or scientifically controversial ones such as fear of cell phone radiation. These controversial emotions could be run through a neosentimentalist filter: We would then have to ask whether they are coherent in the sense mentioned above in Section 4, or can be regarded as secondary properties (as properties which are response-dependent but nevertheless based on dispositional structures of the object itself).

In Section 4, the main neosentimentalist criteria of appropriateness were sketched. If the emotions pass these criteria and are considered appropriate, they still have to be carefully weighed against the costs of the enterprise. In controversial cases, on the level of normative assessment, there will be no easy answer available, and it will require a qualitatively enriched approach to deal with the problems.

In summary, this paper has argued against an antisentimentalist approach in risk studies as it was recently supported by Sjöberg. Such an antisentimentalist approach eliminates emotional factors out of a proper analysis of risk, and aims at an objectivist and rationalist understanding of risk. In the first and second section, I documented the history of interdisciplinary risk studies in sociology and social psychology that primarily opposed an oversimplifying “objectivism of risk.”

In contrast, especially authors in sociology developed a socio-constructivist notion of risk that paid attention to emotional factors in risk perception. After criticizing Sjöberg’s antisentimentalism in the third section, in the fourth section, I tried to shed some light on the metaethical implications of a constructivist view and sketched a neosentimentalist account of risk as an alternative. In a fifth section, I pictured the possible normative implications of a neosentimentalist framework and, on the level of normative ethics, elucidated a recipient-oriented theory of risk assessment.

Acknowledgments I wish to thank Sabine Roeser and Niels Taubert for their helpful comments and Dieter Birnbacher for the opportunity to co-teach a course with him on risk ethics. Niels Taubert’s expertise on the sociology of risk has substantially contributed to the first and second part of this paper. It goes without saying that mistakes are all mine.

References

- Beck, U. 1986. *Risikogesellschaft. Auf dem Weg in eine andere Moderne*. Frankfurt a. M: Suhrkamp.
- Beck, U. 1988. *Gegengifte. Die organisierte Unverantwortlichkeit*. Frankfurt a. M: Suhrkamp.
- Birnbacher, D. 1999. Ethische Dimensionen bei der Bewertung technischer Risiken. In *Technikverantwortung. Güterabwägung – Risikobewertung – Verhaltenskodizes*. Lenk H., and Maring M., eds., 136–147, Frankfurt/M: Suhrkamp.
- Birnbacher, D. 1993. Ethische Fragen der Risikobewertung am Beispiel der Gentechnologie. In *Natur in der Krise. Philosophische Essays zur Naturtheorie und Bioethik*. Löw, R., and Schenk R., eds., 31–51, Hildesheim: Ontos.
- Birnbacher, D. 1996. Risiko und Sicherheit – Philosophische Aspekte. In *Risikoforschung, Disziplinarität und Interdisziplinarität. Von der Illusion der Sicherheit zum Umgang mit Unsicherheit*. Banse, G., ed., 193–210, Berlin: Akademie Verlag.
- Blackburn, S. 2000. *Ruling Passions. A Theory of Practical Reasoning*. Oxford: Oxford University Press.
- D’Arms, J., and J., Daniel. 2000a. Sentiment and value. *Ethics* 110: 722–748.
- D’Arms, J., and J., Daniel. 2000b. The moralistic fallacy: On the appropriateness of emotions. *Philosophy and Phenomenological Research* 61(1): 65–90.
- De Sousa, R. 1990. >The Rationality of Emotions. Cambridge MA: MIT Press.
- De Sousa, R. 2003/2007. *Emotion. Stanford Encyclopedia of Philosophy*. Accessed June 2009. <http://plato.stanford.edu/entries/emotion/>.
- Douglas, M., and A., Wildavski. 1982. *Risk and Culture. An Essay on the Selection of Technological and Environmental Dangers*. Berkeley: University of California Press.
- Ferguson, E., K., Farrell, K. C., Lowe, and V., James. 2001. Perception of Risk of blood transfusion: Knowledge, group-membership and perceived control. *Transfusion Medicine* 11: 129–135.
- Finuncane, M. L., A., Alhakami, P., Slovic, and S. M., Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioural Decision Making* 13: 1–17.
- Fischhoff, B., P., Slovic, S., Lichtenstein, S., Read, and B., Combs. 1978. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences* 9: 127–152.

- Gaskell, G. et al. 1997. Biotechnology and the European public concerted action group, Europe ambivalent on biotechnology. *Nature* 387: 84S5–7.
- Gaskell, G., N., Allum, M., Bauer, J., Jackson, S., Howard, and N., Lindsay 2003. Ambivalent GM nation? Public attitudes to biotechnology in the UK, 1991–2002. *Life Sciences in European Society Report: London School of Economics and Political Sciences*.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings, A Theory of Normative Judgment*. Cambridge MA: Harvard University Press.
- Gibbard, A. 2005. Angemessenheit und Mittelmaß. btl>Wie Gefühle und Handlungen aufeinander abgestimmt werden. In *Adam Smith als Moralphilosoph*. C. Fricke, and H.-P. Schütt eds., 277–303, Berlin: De Gruyter.
- Goldie, P. 2009. *The Emotions. A Philosophical Exploration*. Oxford: Clarendon.
- Graham, G. 2001. Technological danger without stigma: The case of automobile airbags. In *Risk, Media, and Stigma. Understanding Public Challenges to Modern Science and Technology*. Flynn, J., Slovic, P., and Kunreuther, H., eds., 241–256, London: Earthscan.
- Hansson, S.-O. 2004. Ethical criteria of risk acceptance. *Erkenntnis* 59(3): 291–309.
- Hume, D. 1975. *Enquiry concerning human understanding*. In *Enquiries concerning Human Understanding and concerning the Principles of Morals*. 3rd ed., revised by P. H. Nidditch, L. A. Selby-Bigge ed., Oxford: Clarendon Press.
- Hume, D. 1978. *A Treatise of Human Nature*. L. A Selby-Bigge and P.H. Nidditch. Oxford: Oxford University Press.
- IPPNW-Report. 2006. *Gesundheitliche Folgen von Tschernobyl 20 Jahre nach der Reaktorkatastrophe*. Metaanalyse April 2006. Accessed 02/2007. <http://www.ipnw.de/stepone/data/downloads/4e/00/00/Gesundheitliche%20Folgen%20von%20Tschernobyl%20%20Stand%2018April%202006.pdf>.
- Krohn, W., and G., Krücken. 1993. Risiko als Konstruktion und Wirklichkeit. eine Einführung in die sozialwissenschaftliche Risikoforschung. In *Risikante Technologien: Reflexion und Regulation*. Krohn W., and Krücken, G., eds., 9–44, Frankfurt/M: Suhrkamp.
- Lengfelder, E. 1990. *Strahlenwirkung; Strahlenrisiko. Daten Bewertung und Folgerungen aus ärztlicher Sicht*. Landsberg: ecomed.
- Luhmann, N. 1993. Risiko und Gefahr. In *Risikante Technologien: Reflexion und Regulation*. Krohn, W., and Krücken, G., eds., 138–185, Frankfurt/M: Suhrkamp.
- Mackie, J. L. 1991. *Inventing Right and Wrong*. New York: Penguin.
- Manes, A. 1913. *Versicherungswesen*. 2. umgearbeitete und erweiterte Auflage. Leipzig/Berlin: Teubner.
- McDowell, J. 1997. Values and secondary qualities. In *Moral Discourse and Practice: Some Philosophical Approaches*. S. Darwall, A. Gibbard, and P. Railton, eds., 201–213, Oxford: Oxford University Press.
- Nichols, S. 2004. *Sentimental Rules: On the Natural Foundation of Moral Judgment*. New York: Oxford University Press.
- Nohrstedt, S. A. 1991. The information crisis in Sweden after chernobyl. *Media, Culture and Society* 13(4): 477–497.
- Nussbaum, M. 2001. *Upheavals of Thought. The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Otway, H., and P. D., Pahner. 1979. Risk assessment. *Futures*. April 1979, 122–134.
- Perrow, C. 1984. *Normal Accidents: Living with High Risk Technologies*. London: Basic Books.
- Peters, H. 1959. *Die Geschichte der Sozialversicherung*. Bad Godesberg: Asgard.
- Prinz, J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Rescher, N. 1983. *Risk. A philosophical Introduction to the Theory of Risk Evaluation And Management*. New York: Lanham.
- Roeser, S. 2006. The role of emotions in judging the moral acceptability of risk. *Safety Science* 44: 689–700.
- Roeser, S. 2007. Ethical intuitions about risks. *Safety Science Monitor* 3 11: 1–13.

- Roeser, S. 2009. The relation between cognition and affect in moral judgments about risk. In Asveld and Roeser (eds.), *The Ethics of Technological Risk*, London: Earthscan, 182–201.
- Roeser, S. 2010. Emotional reflection about risks. In *Emotions and Risky Technologies*. Roeser, S., ed., Springer, (this volume).
- SSK (Strahlenschutzkommission) (2006) *20 Jahre nach Tschernobyl – Eine Bilanz aus Sicht des Strahlenschutzes. Stellungnahme der Strahlenschutzkommission*. Accessed 02/2007. <http://www.ssk.de/werke/volltext/2006/ssk0603.pdf>.
- Scheler, M. 1980. *Der Formalismus in der Ethik und die materiale Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus*. Bern/ München: Lang.
- Scherer, K. R. 1999. Appraisal theories. In *Handbook of Cognition and Emotion*. T. Dalgleish, and M. Power eds., 637–663, Chichester: Wiley.
- Schwartz, S. H. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 Countries. *Advances in Experimental Social Psychology* 1992: 1–62.
- Schöpfer, G. 1976. *Sozialer Schutz im 16.-18. Jh. Ein Beitrag zur Geschichte der Personenversicherung und der landwirtschaftlichen Versicherung*. Graz: Leykamp.
- Shrader-Frechette, K. 1991. *Risk and Rationality. Philosophical Foundations for Populist Reforms*. Los Angeles: Berkeley.
- Siegrist, M. 2003. Perception of gene technology and food risks: Results of a survey in Switzerland. *Journal of Risk Research* 6: 45–60.
- Sjöberg, L. 2003. Risk perception, emotion and policy: The case of nuclear technology. *European Review, Cambridge University Press* 11(1): 109–128.
- Sjöberg, L. 2006. Will the real meaning of affect please stand up? *Journal of Risk Research* 9(2): 101–108.
- Sjöberg, L., and B.-M., Drottz-Sjöberg. 1997. Physical and managed risk of nuclear waste. Physical and managed risk of nuclear waste. *Risk – Health, Safety & Environment* 8: 115–122.
- Sjöberg, L., and E., Engelberg. 2005. Life styles and risk perception consumer behaviour. *International Review of Sociology* 15(2): 327–362.
- Slovic, P. 1987. Perception of risk. *Science* 236: 280–285.
- Slovic, P. 2000. *The perception of risk*, Earthscan: London.
- Smith, A. 1759/1976. *A Theory of Moral Sentiments*. D. D. Raphael ed., Oxford: Clarendon.
- Solomon, R. C. 2003. *Not Passion's Slave: Emotions and Choice*. Oxford: Oxford UP.
- Starr, C. 1969. Social benefit versus technological risk. What is our society willing to pay for safety? *Science* 165: 1232–1238.
- Steinfath, H. 2001. Gefühle und Werte. *Zeitschrift für philosophische Forschung* 55(2): 30–54.
- Taubert, N. C., and F., Kraemer. 2007. Grenzenloses Experimentieren? Replik auf Wolfgang Krohn 2007: Realexperimente – Modernisierung der ‘offenen Gesellschaft’ durch experimentelle Forschung. *Erwägen, Wissen, Ethik* 18(3): 407–410.
- Townsend, E. March 2006. Affective influences on risk perceptions, and attitudes towards, genetically modified food. *Journal of Risk Research* 9(2): 125–129.
- Wiedemann, P. M., and H., Schütz 2000. Of tales and talks. Using Risk Stories to Understand and overcome different perspectives in risk communication. Presentation at the SRA Conference 2000, Washington DC, cited in. Programmgruppe Mensch Umwelt Technik (MUT), Peter Wiedemann and Anne Brüggemann, Arbeiten zur Risikokommunikation Heft 82, Jülich, Juli 2001, 13–14. Accessed June 2009. www.fz-juelich.de/inb/inb-mut/publikationen/hefte/heft_80.pdf.
- Wiggins, D. 1998. A sensible subjectivism. In David Wiggins, *Needs, Values, and Truth*. 3rd ed., 185–214, Oxford: Oxford University Press.

Risk Emotions and Risk Judgments: Passive Bodily Experience and Active Moral Reasoning in Judgmental Constellations

Mark Coeckelbergh

1 Introduction

Experts typically accuse lay people of “emotional” responses to technological risk as opposed to their own “rational” judgment. When people oppose a particular technology, they are said to be lacking information, scientific education, and rational judgment.

This attitude towards lay people judgments is in tune with risk perception and risk communication research that qualifies lay people’s responses in terms of bias, affect, or feeling (e.g. Slovic et. al. 2004; Keller et al. 2006). Now there is no doubt that emotions play an important role in risk judgments. Paul Slovic and others have done much to show that emotions play a role in risk perception (Slovic et al. 2004; Tversky and Kahneman 1974; see also the overview presented in Peters et al. 2006 and in Covello and Sandman 2001).¹ But this literature has a normative dimension to it as well: it does not only show that emotions play a role; it also communicates an attitude of mistrust towards emotions when it comes to their role in risk judgment.² Consider the concepts used. For instance, emotions are seen as part of a “heuristics”, that is, of judgmental short-cuts. While I do not wish to challenge the results of empirical work on heuristics and biases referred to above, using the terms “bias” and “heuristics” suggests that the authors interpret their results as implying that emotion does not contribute to proper risk judgment. Consider also the concepts “risk as feeling” versus “risk as analysis” (Slovic et al. 2004). Although Slovic has

M. Coeckelbergh (✉)

Philosophy Department, Twente University, Enschede, The Netherlands
e-mail: m.coeckelbergh@utwente.nl

¹The results of these studies are in tune with conclusions from neuroscientists and empirically oriented philosophers who have used neuroimaging technology (fMRI scans) to show that emotions play a role in moral judgment and moral cognition (Greene et al. 2001; Young and Koenigs 2007).

²A similar remark can be made about Greene’s work: Greene observes that emotions are important but, as Roeser has shown, he holds the view that they should not play such an important role since they are reflections of evolutionary formed prejudices (Greene 2003; Roeser 2010).

recognised the limitations of risk science and has argued for recognizing citizens as partners in the exercise of risk assessment (Slovic 1999), the concepts that frame the discussion tend to merely re-enforce the polarisation between laypeople and experts (Coeckelbergh 2009).

In moral theory, this attitude of risk experts is compatible with a view of emotions as irrational forces that should be separated from moral judgment. Reason, not emotions, should guide these judgments. This view is often attributed to Kant, who made a rigid separation between the moral sphere (freedom) and the empirical sphere (determinism). Since, so it is argued, emotions belong to the latter category, they should be excluded from moral judgment. Moreover, on this view it is unthinkable that “moral sense” or “moral sentiment” could be the basis of morality, as the moral sentiment tradition (Hume, Smith, etc.) has it. As Kant has argued in the *Groundwork* and elsewhere, we should seek the basis of morality in reason, not in sentiment (Kant 1785).

If we wish to oppose such views – at least for what concerns their view of emotions – and link emotions to judgment in a stronger way, what are our options? Do we have to understand emotions as judgments, and reject views that link emotions to the body? Or should we recover such a body-oriented view, and reject the cognitivist view? Many contributions to the contemporary discussion about emotions side with one of these camps. In this paper I attempt to steer a different course. I argue that we should neither conflate emotions with judgment nor separate them entirely, but rather provide an account of the exact relation between the two which does justice to the specificity of both, one aspect of which I characterise in terms of activity and passivity. Using Angela Smith’s “rational relations” view and employing the metaphor of a “constellation”, I make a suggestion about how to (re)view the relation between emotions and judgment,³ respond to recent arguments by Sabine Roeser, Jesse Prinz, and Peter Goldie, and explore the implications for discussions about technological risk.

2 Are Emotions Judgments or Bodily Changes? Mind and Body, Activity and Passivity

A straightforward route to give emotions a more important role in relation to judgment than Kantians are willing to and contemporary psychologists of perception unintentionally suggest, is to embrace cognitivism and argue that emotions *are* judgments (Solomon 1980, 2003, 2006), and/or that they are assessable as rational or irrational (de Sousa 1987).

³Note that my account does not make the Kantian distinction between moral and prudential judgment. To do so would imply that emotions can play a role in prudential judgment but should not “interfere” in moral judgment, a view which I reject. The account developed in this paper is applicable to both moral and prudential judgments and hence to all risk judgments in so far as they involve such judgments.

In *The Rationality of Emotion* (1987) Ronald de Sousa has argued that while emotions *are* not beliefs (de Sousa 1987, p. 173) and are often experienced as “gut feelings” (198), they can be assessed as rational or irrational since they are a kind of perception, “apprehensions of real properties in the world” (201). De Sousa tells us that we have “emotional repertoires” (236) that frame our “possibilities of experience” (332). This view does not imply that our emotions are determined by our previous experiences and our nature; De Sousa recognises that we can “regealt” our paradigms and have some control (263).

Interestingly, de Sousa explicitly recognises the antinomy of activity and passivity as one of the philosophical problems emotions lead us to. He writes: “The word “passion” suggests passivity; yet in many ways emotions seem to express our most active self.” (de Sousa 1987, p. 2) and are “sometimes the very embodiment of the will” (de Sousa 1987, p. 46). How shall we understand the latter claim? Robert Solomon, another cognitivist, writes in his book *Not Passion’s Slave*:

Emotions are not occurrences and do not happen to us. I would like to suggest that emotions are rational and purposive rather than irrational and disruptive, are very much like actions, and that we choose an emotion much as we choose a course of action (Solomon 2003, p. 3).

In tune with his earlier work (Solomon 1980), Solomon thinks of emotions as active and argues that therefore we are responsible for them.⁴ He provides the following example. If I am angry at someone for stealing my car, then this is a judgment since I believe that I have been wronged: “If I do not believe that I have somehow been wronged, I cannot be angry” (Solomon 2003, p. 8). He concludes that “emotion is a normative judgment, perhaps even a moral judgment” (8). This does not mean that we are always aware of making such a judgment, making such a choice; they are “hasty and typically dogmatic judgments” and in this sense they are “blind” (17). Emotional judgments are “spontaneous” and “typically not deliberative” (96). Nevertheless, emotions can be rational (or not) in the same way as judgments can be rational (or not) (11, 35). And since judgments are actions, Solomon argues, emotions too are actions⁵: they are “aimed at changing the world” (11).⁶

But what about the passive side of emotional experience? What about feelings? Solomon discusses this question at length in the last chapter of the book (“On the

⁴ A similar view can be found in Sartre’s *The Emotions* (Sartre 1948).

⁵ Note that recently emotions have also received a more prominent place in the philosophy of action, which seems to support the cognitivist view. In their article “Emotion and Action” Zhu and Thagard argue against the view that emotions are irrational and that they “merely happen to people” (Zhu and Thagard 2002, p. 19). Drawing on research in cognitive neuroscience, they conclude that “emotions contribute significantly to the processes of action generation as well as action execution and control” (34).

⁶ This sounds like Sartre, but Solomon rejects Sartre’s view that emotions have a “magical” function. He calls wanting to undo the past, stereotype responses, avoiding unusual situations etc. “pathological ways of choosing our emotions” (Solomon 2003, 13). According to Solomon, emotions do not merely change our *view* of the world, the also (make us) change the world. Note also that the cognitivist is similar to the Stoic view of emotions, as Martha Nussbaum (2001 and Miriam van Reijen have shown (van Reijen 1995, 2005).

Passivity of the Passions”). He argues that not all of our emotions are suddenly provoked or “provoked by sudden circumstances” (Solomon 2003, p. 201). *Some* are (226). Often there is a set of events, many emotions are “enduring processes” and last a long time (203). Thus, what he calls the “emergency paradigm” (203) is not the only way to understand emotions. Moreover, Solomon argues against the equation between emotion and feelings. His argument is “the simple fact that we often have an emotion without experiencing any particular feeling” (31). For him, emotions are a way of seeing and experiencing (75) rather than feeling. Solomon does not deny the existence of feelings, of “being in a passion” and “becoming emotional” – he even claims that “emotional judgments are “dispassionate” only in pathological circumstances” (109), but argues that feelings are not the emotion (30). Emotions, he thinks, are rational and have a “logic” of their own (35). They are “a lot like thoughts” (206). Against James (see below), Solomon defends a cognitivist view of emotion that he expresses as follows: “An emotion is a system of concepts, beliefs, attitudes, and desires” (87). For Solomon, talk of passivity is “misleading” (212). He claims that even panic and rage involve cognition and judgment (214). He refers to the Stoics to argue that what is passive about emotions is “nothing more than an indicator about what we ourselves were actively doing, how we were living” (216). Finally, Solomon asks whether or not the *expression* of emotion may be involuntary and connects even expression with responsibility: we are responsible for “the emotion *as* expression” – the two cannot be separated since expression cannot be stripped-down to mere bodily movement (222). Generally, Solomon connects this point about responsibility for our emotions with the following normative ideal:

Arguing as I have amounts to nothing less than insisting that we think of ourselves as adults instead of children, who are indeed the passive victims of their passions. (Solomon 2003, p. 232)

In other words, Solomon’s descriptive view is connected with the normative view that we *should* take responsibility for our emotions. Moreover, with regard to *moral* judgment, Solomon argues that emotional judgments are evaluative and involve “cognition, appraisal, and evaluation” (100).

Thus, both de Sousa and Solomon see emotions as active, cognitive, and evaluative elements for which we are responsible.

However, although these accounts succeed in showing the importance of emotions in relation to moral judgment, they do not sufficiently account for the “raw”, bodily and passive aspect of much emotional experience. In response to those who accused him of neglecting the body, Solomon opened up the category of judgment to bodily changes: he interpreted such changes as judgments, using the term “judgments of the body” (Solomon 2003, p. 191). But this unhelpfully inflates the meaning of judgment. As Matthew Ratcliffe puts it in his book review, “it might as well incorporate everything” (Ratcliffe 2003). In particular, it is hard to see how “judgments of the body” can be called *moral* at all (or prudential, for that matter). Furthermore, although de Sousa stresses that emotional rationality is not the same as rationality of belief or desire (de Sousa 1987), it seems impossible to reconcile

his view with the passive and bodily side of emotional experience. If anger is experienced by a person as something that happens to him or her, it seems odd to assess the rationality or irrationality of the anger itself (as opposed to beliefs or actions that are connected with it). Of course these beliefs or actions might be justified. But anger, as far as the bodily experience of it is concerned, appears (to the person and those involved) as rational or irrational as natural occurrences.

To account for the bodily side of emotional experience, then, we may want to turn to William James's early view of emotions as the experience of bodily changes (James 1884) and Jesse Prinz's version of this view (Prinz 2004b). This turn has happened and is happening, and, as Heleen Pott has argued, is partly motivated by trying to cope with the problem of the passivity of emotions (Pott 2005, p. 117). People such as Michael Stocker and David Pugmire have put this problem on the agenda. Cognitivist theory could not and cannot sufficiently account for the often involuntary, uncontrollable, and obsessive character of emotional experience (Pott 2005, p. 118).

In his famous article "What is an Emotion?" (1884) William James argues that emotion is not something mental that produces a bodily expression, but is to be equated with the feeling of bodily changes:

Our natural way of thinking about these standard emotions is that the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression. My thesis on the contrary is that *the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion.* (James 1884, pp. 190–91; his emphasis)

In his chapter "The Emotions" in *The Principles of Psychology* this definition is repeated (James 1890, p. 1065). James's view that there is an "immediate physical influence" (James 1884, p. 197) explains the passive side of emotional experience. Sometimes emotions follow a kind of short-cut, bypassing cognitive processes. Inspired by James, Jesse Prinz has argued that emotions are embodied appraisals, "gut reactions" (Prinz 2004b). In addition, he recently has embraced sentimentalism, which says that "to believe that something is morally wrong (right) is to have a sentiment of disapprobation (approbation) towards it" (Prinz 2006, p. 33). He connects this to James and his own "embodied appraisal" view in the following way.

Moral judgments express sentiments, and sentiments refer to the property of causing certain reactions in us. The reactions in question are emotions, which I regard as feelings of patterned bodily changes. (Prinz 2006, p. 34)

Thus, James and Prinz offer us a view of emotions that accounts for their bodily side, but at the cost of downplaying the role of practical rationality. Although cognitivist theories have put too much emphasis on rationality and action, in this account this dimension seems entirely absent. However, Prinz also adapts James. He criticizes James for his "failure to reckon with what can broadly be regarded as the rationality of emotions" and proposes to see emotions not only as somatic but also as "semantic: meaningful commodities in our mental economies" (Prinz 2004a, p. 45). What does this mean? Prinz defines an appraisal as "any representation of an organism-environment relation that bears on well-being" (57). For Prinz, emotions are (only)

like evaluative judgments: “they represent roughly the same thing that evaluative judgments present, but they do it by figuring into the right causal relations, not by deploying concepts or providing descriptions. Our perceptions of the body tell us about our organs and limbs, but they also carry information about how we are faring” (57).

Now this seems to be a purely causal understanding of emotion that has little or nothing to do with moral judgment. In such a view, emotions are like warning lights on a dashboard: they let us know how we are faring. Moral judgment is supposed to be more (complex) than that, especially judgment concerning technological risks is not simply a problem of the relation between organism and environment. And on James’s view, the relation between emotion and judgment becomes one of expression. Thus, judgment as active moral reasoning is either reduced to emotion and therefore no longer moral reasoning and no longer active at all, or it is completely separated from emotion. To refer to James’s example: my judgment that the bear is dangerous is only an expression of bodily changes or if it is “separate” from these changes it is irrelevant and superfluous since the body has already “judged” or “made up its body” (as opposed to “made up my mind”). Both the cognitivist and the Jamesian view tie up emotions and judgment so closely that we can no longer make sense of the folk psychology idea of an emotion and a judgment as two different things. Some may see that as progress, but I would like to try harder to save it, without having to endorse what I have called the Kantian and psychology of risk perception view. Can we steer a middle course between separation (Kant and risk perception view) and identity (cognitivism and James)? Can we describe the relation between emotion and judgment in a way that avoids these two extremes?

3 Emotions and Rational Relations

To further reflect on the relation between emotions and judgment, I seek inspiration from Angela Smith’s “rational relations” view. In her paper “Responsibility for Attitudes: Activity and Passivity in Mental Life” (2005), Smith argues that there is a normative connection between spontaneous attitudinal reactions and our underlying evaluative judgments and commitments (Smith 2005). Let me briefly explain her position. Against what she calls the volitional view of responsibility, which holds that choice and voluntary control is a necessary condition for responsibility (Smith 2005, 238), she points to “different ways in which our attitudes and reactions can be said to reflect our evaluative judgments” and argues that these connections are sufficient for responsibility (Smith 2005, p. 237). However, these “reactions” are not the “gut reactions” or “bodily changes” Prinz and James have in mind. Smith makes a distinction between “brute sensations, which simply assail us” and “spontaneous reactions” which “reveal, in a direct and sometimes distressing way, the underlying evaluative commitments shaping our responses to the situations in which we find ourselves” (Smith 2005, p. 250). Both “do not arise from conscious choice or decision” (Smith 2005, p. 250), but the spontaneous reactions are rationally connected to our evaluative commitments and are therefore morally relevant. She argues that

physical reactions can serve as “moral indicators of our evaluative judgments”, but that this relation is “purely causal”, whereas our attitudes “are not merely the causal effects of our judgment” but “active states, in the sense that they essentially involve our judgmental activity” and therefore we are responsible for them (Smith 2005, p. 258). The latter are “judgment dependent” whereas the former are not.

Let me now apply this “rational relations” view to the discussion about the relation between emotions and judgment. Are emotions like the attitudes and the spontaneous reactions Smith describes? Or are they more like “brute sensations”?

We need not deny that there are emotional experiences that can be described as “brute sensations”, as “bodily changes”, and as “gut reactions”. With regard to such experiences, we can study the causal relations, discover how the body works. But the emotional experiences that are relevant *morally* are usually not of this kind. Consider the emotional experiences relevant to risk judgment. For example, in response to a picture of people who died from an atomic bomb explosion, we may feel direct, causal fear of death, but the morally relevant fear is our fear of the technology being used for this purpose. This fear is related to the “raw” fear of death, but it is much more: it is part of a moral judgment which also involves other elements. When it comes to emotions that are morally relevant, my thesis is that these emotions – however passive, bodily, overwhelming etc. – are rationally connected to judgments and the values on which we based them, or are at least always open to such a connection. Thus, by distinguishing between two kinds of emotional experience, I am able to account for both the passive, bodily side and the active, reasoning side. Emotions are not judgments, as cognitivists claim, but they can be rationally related to judgment (or not), and this *relation* is open to assessment. Emotions are strongly related to action, but the connection need not be seen in causal terms. And James and Prinz are right to call attention to the bodily and passive side of emotional experience, but they fail to distinguish between emotions as “gut reactions” and emotions as a mental state similar to attitudes that can be rationally connected to our normative commitments. The latter have a passive side as well, but this passivity is not best described by reference to causal processes in the body, but by reference to the normative, evaluative commitments we have, the values and beliefs we hold. Such commitments bring us in a position that Harry Frankfurt has described with Luther’s phrase “I can do no other” (Frankfurt 1988, 1999). However, in contrast to Frankfurt, we need not understand this as being necessarily a question of “voluntary necessity”. It can also be a case of “rational” necessity: “I can feel no other” since (for example) I care about X. There is a rational connection between my care for X and my emotion, and this connection renders me in a situation of passivity and “necessity”.

The term “necessity” is perhaps not the most adequate term since it might be taken to refer to gut reactions that have a causal connection. The emotional experience considered here must instead be located in the sphere of freedom and morality.⁷ By making such a distinction between freedom and necessity, I arrive at Kant again,

⁷ Note that our emotion and judgment can also be non-moral, but here I consider (risk) judgment as moral judgment.

perhaps, but without buying his denigrating view of emotions. Emotional experience, now, can be understood as partly belonging to the sphere of necessity when it is re-action (think of the bear), but it must also be seen as *possibly* having its home in the sphere of freedom, in cases when it is rationally related to our deepest moral commitments. In such cases, our action taken on the basis of the emotion can rightly be called free, since it is *my* commitment and *my* judgment which is rationally related to *my* emotion. At the same time, we must also recognise the deep passivity involved in the emotional experience, a passivity that can be felt, bodily.

4 Emotions as Part of a Judgmental Constellation

If this view is right, does it imply that emotions can be only consequences of evaluative commitments? Or can they also change those commitments? Let me say more about the nature of the “rational relation” between emotions and judgment. Angela Smith argues that if fear involves the judgment that something is dangerous, there is not just a causal connection but a “conceptual connection” or a “rational connection” (Smith 2005, p. 270). But her examples mainly suggest a one-way process: there is a rational connection in the sense that our evaluative commitments give rise to attitudes or emotions. This view remains too close to views that see emotions merely as expressive. In these views emotions are not given their full significance as experiences that themselves can give rise to further reasoning and judgment, perhaps even change of our evaluative commitments. Rather than conceptualising this connection in terms of “expression” or “reflection”, I propose to understand the relation between emotions and judgment in the following terms: emotions are part of a “constellation of moral judgment”. I coin this concept in order to avoid both identity and separation, and to allow a two-way process. A constellation is a group of “some-things” and a group allows for mutual relations. In astronomy, stars form groups as they are related by way of gravity, but they are separate entities. Similarly, emotions are not judgments. The two are not identical, but they are related. A large part of the emotional experience that is morally relevant in risk judgments and other judgments is not of the “bear” kind, not of the re-active kind, but stands under the influence of the gravity field constituted by evaluative commitments, values, beliefs, and concerns, and vice versa.⁸ Together, these elements form a judgmental constellation, a group of elements that together constitute moral judgment.

Let me give an example in the context of judgments concerning technological risk. Many lay people, while benefiting from nuclear technology through the consumption of energy produced by it, reject the technology. Is this irrational? Is it a typical case of bias? Is it an “emotional” reaction? Perhaps it is a matter of emotions, but not of emotions alone and not of emotions as mere “gut reactions”. Of course imagery of death and destruction may strike us with horror in

⁸ Note that I have a more pluralist definition than Roberts (see also his contribution to this volume): the value dimension enters not only via concern.

a very direct, re-active way. But most part of the emotional experience involved here – which can also be quite strong, be felt physically, and testify of deep passivity – may be rationally related to strong moral convictions and commitments, to knowledge and belief, etc. For example, imagine that people believe that there is a problem with potential transfer of technology from the context of energy production to that of warfare and that there is a problem with nuclear waste. Furthermore, imagine that they are committed to peace, and they care for future generations. If their emotions concerning nuclear technology are rationally related to these beliefs and commitments, then these emotions can adequately be regarded as forming a part of, and co-constituting, a judgment, a moral judgment concerning the technology.

For discussions about technological risk, this implies that we (experts and others) should apply a principle of charity or openness when we meet other people's expressed views that involve emotion: we should assume that they behave as the rational and emotional beings that they are, and that their emotions are part of a judgmental constellation.

This argument can be seen as an interpretation of, or an addition to, the "ideal speech situation" requirements Habermas has argued for, conditions which are meant to enable what he calls "free discourse" and "communicative action" (Habermas 1981). It also refers to Kant's idea that we are, or should be, moral equals.

Of course, as with all judgments, questions can be asked with regard to the quality and/or rightness of the judgment. To say that emotions should be taken seriously since there is a good chance that they are part of a judgmental constellation, is not to say that all judgments made in a particular case with regard to a particular technology are right. For example, commitments may be morally problematic, and the way the various elements are related may show some irrationality. In the best case, this can be clarified in further discussion. But the emotions themselves are not rational or irrational. The rationality is not atomistic but holistic here, in the sense that the rationality of the constellational relations as a whole must be assessed in order to say something about the quality of the overall moral judgment. Finally, it may be that there are limits to the degree to which we can assess a moral judgment as right or wrong, or in terms of its rationality. And even if it would be possible to a large extent, it need not mean that a judgment is absolutely right or absolutely wrong: between these extremes (if they exist at all) lies a universe of moral possibilities.

5 Comparison to Recent Interpretations of Cognitivism and James

In order to further clarify my view, let me compare it with three recent contributions to the philosophy of emotions that seek to adapt cognitivism and James towards a "middle way" between both extremes.

5.1 Roeser's Cognitivism

According to Roeser, we need emotions in order to judge the moral acceptability of technological risks (Roeser 2006). She justifies this claim by arguing that emotions “have cognitive and affective aspects at the same time” (Roeser 2010, p. 11). She gives examples such as feelings of indignation, shame, and guilt. Thus, she subscribes to the cognitivist idea that emotions are value judgments, but in contrast to standard cognitivism she argues that they are feelings at the same time – thereby incorporating the Jamesian insistence on the affective aspect of emotions. Drawing on intuitionism and arguing against Greene she claims that we need emotions in order to have access to objective moral truths: they can be “a form of judgment and insight into objective moral truths” (21).

This view stays too close to the cognitivist view that equates emotions with judgments. When we feel indignation, this feeling is not itself a judgment; it is more adequate to understand what goes on by saying that there can be a rational relation between the expression of emotion and the moral judgment. Roeser's position that emotions are a form of judgment and can provide insight into objective moral truths seems to imply that emotions can be right or wrong. But this does not correspond well to some of our intuitions. For example, when we say (perhaps following Aristotle) that emotions can blind judgment and make a debate impossible, we distinguish between emotions and judgment. If we equate emotions with judgment, as Roeser does, we have no way to account for this experience. Roeser could respond that this is a case of a failed, inadequate perception of moral truth. But such a description does little justice to the adversarial, hostile relation between emotion and judgment experienced in these cases. A rational relations view can do a better job. If there is no rational relation between our emotion and the rest of our judgmental constellation, then it is appropriate to either say that there is an internal struggle between the emotion and the other elements of the constellation or say that “emotion blinds judgment”: while the other elements of our judgment are intact, emotion (temporarily) blocks off our judgment, renders the other elements of the judgmental constellation mute. If, on the contrary, our emotion is rationally connected with the rest of our judgment, the emotion enables us to see the world in the light of our judgment, provides a window for judgment, and offers – to others – a royal entry into our judgmental constellation. This enables mutual understanding – although it does not guarantee a consensus. In contrast to Roeser, I do not need to make a claim about the objectivity or rightness of the judgment. My account only claims that emotions can, or cannot, be rationally related to other elements in a judgmental constellation.

Thus, if there is no rational relation, emotion can be seen as separate from the other elements. In this case some may call the emotion “irrational”, but in my view it is not the emotion that is to be described as irrational; it is the rational relation that is missing. The relation is to be evaluated, not only the separate elements.

Are emotions necessary for (a right) moral judgment? One might argue that emotions are a necessary part of a judgmental constellation or that the judgment is not *complete* without the “right” emotion – right in the sense of being rationally related to the other elements in the judgmental constellation. But completeness is neither

necessary nor sufficient for the rightness of a moral judgment. Someone can have a well-built judgmental constellation but the judgment can be wrong. And the judgment can be right, but the emotional expression may be missing or the judgmental constellation may lack completeness if there is no rational relation between emotion and the rest of the constellation. Another response is to say that emotions are not a necessary part of a judgmental constellation, but that its completeness depends on there being a rational relation between emotion and the rest of the constellation. If there is a rational relation, the constellation is more complete. However, if completeness is not related to the rightness of the moral judgment, why care about it at all?

Although completeness is not necessary for the *rightness* of the judgment, it enhances the *quality* of the moral judgment. A more complete judgment is not more *right* but *better*. This is not only so since it employs the capacities we have as cognitive and as affective beings but also since actions following a complete judgment have a broader motivational basis: if our judgment is emotionally robust, we will also feel like acting upon it. Moreover, an emotionally complete judgment assists the agent in developing and maintaining a harmonious mental life and has communicative and rhetorical advantages as compared with the “merely” right judgment. Someone who defends a “right” judgment without being emotionally committed to it will have more problems trying to convince others of his judgment in a discussion than the one who lines up his emotion with his beliefs, commitments, and other elements that are part of his judgmental constellation. Completeness makes the person’s judgment more transparent to others (and to himself) and renders it more convincing.⁹

One could object that looking at emotions this way is to neglect their passive and bodily aspect. If my emotions are drawn into the judgmental sphere, it may seem that they are part of active moral reasoning, and although they retain their bodily aspect, they appear to become the mere expression of judgment. But this impression is misguided. First, the advantage of the rational relations view is that relations can be considered in two directions. Sometimes emotions are the expressions of our beliefs, commitments, ideals, values, etc., but sometimes they *change* our beliefs and commitments. They can change not only our view of the world, of our partner, of our life, etc. but also our commitments, values, and other judgmental elements. As I said, the relations within a judgmental constellation are not one-way; emotions are not only the result of other elements but can also change these elements. Second, if this happens, it can have a strong passive aspect. Consider again Frankfurt’s point that sometimes what we care about puts us in a situation of passivity (Frankfurt 1999). Commitments can also put us in such a situation: sometimes we cannot but have a particular emotion on the basis of a particular deep commitment. However,

⁹ Note that there are more criteria to evaluate the completeness and the quality of a judgment. In his paper “Emotions and Judgments about Risk” Roberts proposes a number of epistemic criteria that can be used to evaluate judgments. For example, a judgment is better if it is epistemically justified and if the subject understands the judgment. Both are unrelated to what Roberts calls the “truth” of a judgment (see Roberts in this volume).

this relation must also be considered in the other direction: if we have a strong emotional, bodily experience in a particular situation, this may change our beliefs, values, commitments. For instance, if someone first holds the belief that a certain job is not really “her thing”, but upon doing the job finds out that she likes it and feels good doing it, then that person is likely to change her beliefs and related cognitive elements.¹⁰

These possibilities for change are good news for moral discussions: if neither our emotions nor the other elements that play a role in our judgment are entirely fixed, they are open to change as a result of communicative processes, which is essential to keep open the practical possibilities for reaching agreement. This assists the process I take Slovic and others to aim for with regard to risk judgment: a collective, communicative process aimed at consensus in which both lay people and experts participate as partners (Slovic 1999). It gives people the opportunity not only to integrate their own judgmental constellations (e.g. by testing their coherence by confronting them with other judgments) but also to build a collective constellation that owes its robustness, harmony, and stability to the fact that both cognitive and emotional elements are shared between the partners and rationally related to one another.

This two-directional view can accommodate the cognitivist intuition that emotion teach us something about what we value, without equating emotion and value judgment. If there is a rational relation between what we feel and what we value as a result of our judgmental constellation, our emotion indeed shows what we value, and, as I said, it can even change what we value by influencing other elements in the constellation.

Note also that this view renders it unnecessary and incorrect to distinguish, as cognitivists and many other emotions theorists generally do, between so-called “moral emotions” (e.g. guilt) and “non-moral emotions” (e.g. fear). On my view, *any* emotion can stand, or not stand, in a rational relation to morally relevant elements of a judgmental constellation. To take an example from risk discussions: our fear of a particular technology may be rationally related to the value we attach to our lives and those of others.

5.2 Prinz’s Adaptation of James

My presentation of Prinz’s view so far was a little unfair, since I understood him as holding a view very close to James. But in this book *Gut Reactions* (2004b) and in his recent paper “Was William James Right About Emotions?” (2007) Prinz defends a modified Jamesian view of emotions that leaves room for cognitive and

¹⁰The situation here is similar to what psychologists call “cognitive dissonance”, except that here the constellation contains emotions as well as cognitive elements. We could call it “emotional-cognitive dissonance”. And similar to the usual response to cognitive dissonance, it is likely that the person will change her beliefs.

judgmental aspects of emotional experience. The Jamesian view of emotion can be summarized as follows:

event ► perception ► bodily change ► emotion

To use Prinz's example: I perceive a snake, my body reacts, and I feel fear. So far, cognition is not involved or only minimally – in so far as perception is a cognitive activity. There is certainly no place for anything like “moral judgment” in the usual sense of the word. But Prinz modifies this scheme by opening it up for cognitive assessment:

event ► perception and/or cognitive assessment ► bodily change ► emotion (Prinz 2007)

Perception can bypass cognition, but there can be also plenty of room for (conscious, deliberate) cognitive assessment. For example, when my expectations are violated, this may also cause a bodily change and an emotion. Emotions are not just a matter of stimulus-response; there is space for cognitive operations.

Compared to the cognitivist view, this model manages to keep judgment and emotion separate (but related). If we replace “cognitive assessment” with “judgment” we get the following model:

(expectations ►) event ► perception and/or judgment ► bodily change ► emotion

Compared to my view, there is a crucial difference: this model describes causal relations. Judgment is related to emotion, but the relation is purely causal. A rational relations view is not necessarily in contradiction with a causal model: it need not deny that there are these causal relations. But it is another way of making sense of emotional and moral experience.

Note that a person need not be aware of the rational relations. Compare this perspective with what happens when philosophers ascribe “reasons for action” to persons: they may “have” good reasons to act in a certain way without explicitly knowing that they have them, without describing what they do and deciding in this way.

Note also that Prinz's causal view can be enhanced by considering the causal chain in *the other* direction:

emotion ► perception and judgment

Emotions can make us see the world differently. But this does not happen in a magical way, as Sartre argued (Sartre 1948). There are causal and rational relations between emotions and the way we perceive and judge.

5.3 Goldie Beyond Cognitivism and James

Peter Goldie's “middle way” between cognitivism and James starts from the claim that emotions are intentional – they are directed towards objects, e.g. I fear *something* (see also Solomon 2003) – but that feelings should not be left out of the picture

(Goldie 2000, p. 4). He understands emotions as involving various elements, including “perceptions, thoughts, and feelings of various kinds, and bodily changes of various kinds” (12). So whereas I open up the term “judgment” to a number of elements, Goldie does so for emotions. This reminds of cognitivism, except that feelings and bodily changes get a place at well. But what place exactly? Looking at Goldie’s account of the intentionality of emotions, it first appears that there is little room for the passive aspects of emotional experience. Goldie speaks of “feeling towards” (19), which means “*thinking of* with feeling, so that your emotional feelings are directed towards the object of your thought” (19; Goldie’s emphasis). The emphasis is on active cognition rather than feeling. However, Goldie sees no tension between the bodily and the intentional. Against mind-body dualism, Goldie argues that “our entire mind and body is engaged in the emotional experience” (55). My view is compatible with such an anti-dualist position since it brings emotions as passive bodily experience together with cognitive elements and other elements of active moral reasoning in one judgmental constellation. Moreover, as I read Goldie he overcomes the activity/passivity dualism, which is interesting given my own purpose in this paper. It is not the case that thoughts are active whereas feelings are passive. Goldie reminds us that passivity is at the “heart” of the cognitive side as well: “the emotions are *passions*: your thoughts and feelings are not always as much under your control as you would want them to be” (58).¹¹ For instance, fear turns otherwise harmless features of the world into dangerous and threatening elements. Put in the terminology I proposed, it means that a particular judgmental constellation, its relations, and its elements are not necessarily under our complete control. Not only emotions, but other elements as well are potentially wild horses – to use a Platonic metaphor. This is a further good reason to reject views that tend to stigmatise emotions as “irrational” as opposed to cognitive elements that are supposed to be “rational”.

To conclude, I doubt whether “the reasons of the heart” are *always* “perfectly intelligible”, as Goldie seems to suggest at the end of this book (Goldie 2000, p. 241), but they often are intelligible, there frequently are rational relations to be discovered and to be understood. There is no reason to presume otherwise when we meet people, their emotions, and their judgments. Rather, we should apply a “principle of charity” and seek out the rational relations between elements in their judgmental constellations.¹²

¹¹ Note that, apart from elements *within* the judgmental constellation, Goldie suggests that there is a sense in which our *actions* themselves can reveal a “passivity”. In his analysis of jealousy he writes: “The passionateness of jealousy is revealed not only in its aetiology and in the way jealous thoughts and feelings can be out of our reasoned control. It can also be revealed in our actions. We can, so to speak, *find ourselves* doing things (. . .).” (Goldie 2000, p. 231) This suggests that we should not only look into the relation between these emotions and the other elements in our judgmental constellation, but also to the relations between elements of that constellation (for instance emotions) and our actions.

¹² If Goldie is right about the passivity of actions (see the previous footnote), we might also want to seek out the rational relations between judgmental constellations and actions.

6 Back to Risk

How can we apply the view I developed to risk? I have argued that emotions are not *necessary* for judgment or for the rightness of a judgmental constellation, and therefore I disagree with Roeser's claim that "emotions are an indispensable normative guide in judging the moral acceptability of technological risks" (Roeser 2006). However, while they are not indispensable with regard to the judgment's rightness, they enhance the quality of our risk judgments, and this gives us a good reason to take them seriously in discussions about technological risk. Furthermore, if we want to evaluate someone's risk judgment, we should not only look at the risk emotions themselves (or their verbal or bodily expression), but ask about their potential relation with that person's beliefs, commitments, and other cognitive elements.

One difficulty that arises when we turn to risk, however, is that the discussion above and the rational relations view from which it draws some inspiration, is made on the individual level. But moral judgments concerning risk and technology (1) do not happen in a vacuum but in a social and cultural context, (2) are matters of social and political concern and therefore are, or should be, part of the public sphere, and (3) are sometimes collective judgments. However, this need not be a problem for the account presented here. It implies a direct relation between public and private in the sense that the elements that make up the judgmental constellation – beliefs, emotions, commitments – are potentially shared and public. This makes public communication and discussion about the moral aspects of risk possible.

These insights can be used to critically assess the concepts used in the psychology of risk perception. Here I will limit myself to the term "bias" and the "perception-judgment" dichotomy. As I said in my introduction, both terms are used in the psychology of risk perception, and tend to have the implication of increasing polarisation by opposing the expert "judgment" and the "facts" to the "bias" and "perception" of the public (see also Coeckelbergh 2009). To avoid this, I propose to redefine the terms in the following way.

First, I propose to understand the term "bias" not as the opposite of a fixed external truth only accessible by experts, but as the description of an emotion that may or may not be rationally related to beliefs, commitments, and other cognitive elements. In this way, we can make sense of the view that both experts and lay people can have emotions, can make judgments, and can have emotions play a role in these judgments. As I argued above, the existence of a rational relation between emotion and other judgmental elements is neither necessary nor sufficient for a judgment to be *right*, but to consider the relation between emotions and judgment in way proposed above is a good route to taking seriously risk judgments by lay people. Before rejecting so-called "emotional" responses as irrational, experts and others must inquire into the possible rational relations between these (expressions of) emotions and the other elements of the judgment of their dialogue partners. In this way, people are regarded and treated as rational (and emotional) beings.

Second, the discussion above allows us to bridge the gap between "perception", ascribed to lay people, and "judgment", ascribed to experts. It has been shown that both elements are related. The possible relations can be approached from two angles.

First, we can describe the *causal* relations. This can be done, for example, by creating a theory that integrates Slovic's two systems – the system of feeling and the system of analysis (Slovic et al. 2004). We might also want to turn to Prinz's model, which links perception and judgment as potential partners at the same stage in the causal chain. In this way, perception and judgment are indirectly connected by emotion. Second, we can link perception and judgment by considering the possibility of *rational* relations between them. This requires a conceptual framework that departs from Slovic and Prinz. On the view defended here, perception and emotions can be part of a judgmental constellation. On both views (the causal and the relational), the opposition between perception and judgment as found in the risk literature can rightly be called a form of bias: it is not rationally related to the insights about emotions and judgment we gained in this discussion.

7 Conclusion: Emotions and Moral Risk Judgments

If the view of the relation between emotion and judgment sketched here is plausible, we must take seriously people's so-called "gut reactions" to technological risk as being potentially rationally related to, but not identical to, judgment. Sometimes they may be "gut reactions" indeed, merely causal reactions to the sight of horror. But more often they are emotions that are rationally related to evaluative elements which, together with the emotions, constitute what can be rightly called a moral risk judgment.

In this view, emotions are not themselves understood as cognitive elements. It recognises the passivity and the bodily aspects of emotional experience. However, it also regards both experts and non-experts as moral, emotional, and rational beings, who have the possibility and a duty to take up responsibility for their emotions, attitudes, and judgments. On this basis, they can communicate and discuss with one another.

An important condition for such a process is the possibility of changing judgments and the willingness to do so. Within the framework of risk judgment understood as a conversation between moral equals, emotions and judgments can change, perhaps *must* change if we are to move forward in discussions about risk. Sometimes this can happen to us. Indeed, the view proposed here recognises the possibility of actively changing our attitude to risk if we judge that there are good reasons to do so, but appreciates that if sometimes technological risk strikes us with fear and horror, that experience teaches us much about what we judge to be important.

Bibliography

- Coeckelbergh, M. 2009. Risk and public imagination: Mediated risk perception as imaginative moral judgment. In *The Ethics of Risk*. L. Asveld, and S. Roeser, eds., London: Earthscan Publishers.

- Covello, V., and P. M., Sandman. 2001. Risk communication: Evolution and revolution. In *Solutions to an Environment in Peril*. A. Wolbarst ed., 164–178, Baltimore: John Hopkins University Press.
- de Sousa, R. 1987. *The Rationality of Emotion*. Cambridge, MA/London: MIT Press.
- Frankfurt, H. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- Frankfurt, H. 1999. *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- Goldie, P. 2000. *The Emotions*. Oxford/New York: Oxford University Press.
- Greene, J. D., R. B., Sommerville, L. E., Nystrom, J. M., Darley, and J. D., Cohen. September 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 14.
- Greene, J. D. 2003. From neutral “is” to moral “ought”: What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience* 4: 847–850.
- Habermas, J. 1981. *Theorie des Kommunikativen Handelns* (Band I + II). Frankfurt am Main: Suhrkamp Verlag. Translated by T. McCarthy (1984) *The Theory of Communicative Action*. Vol. 1 + 2, Boston, MA: Beacon Press.
- James, W. 1884. What is an emotion? *Mind* 9: 188–205.
- James, W. 1890. *The Principles of Psychology*. Vol. II. Cambridge, MA/London: Harvard University Press, 1981.
- Kant, I. 1785. *Grundlegung zur Metaphysik der Sitten*. New York: Routledge, 1991. Translated by H. J. Paton *Groundwork of the Metaphysic of Morals*.
- Keller, C., M., Siegrist, and H., Gutscher. 2006. The role of affect and availability heuristics in risk communication. *Risk Analysis* 26(3): 631–639.
- Nussbaum, M. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Peters, E., K. D., McCaul, M., Stefanek, and W., Nelson. 2006. A heuristics approach to understanding cancer risk perception: Contributions from judgment and decision-making research. *Annals of Behavioural Medicine* 31(1): 45–52.
- Pott, H. 2005. Van James naar Damasio: Balans van dertig jaar emotietheorie. In *Emoties: Van stoïcijnse apatheia tot heftige liefde*. M. van Reijen, ed., Kampen: Klement.
- Prinz, J. 2004a. Embodied emotions. In *Thinking about Feeling*. R. C. Solomon, ed., 44–59, Oxford/New York: Oxford University Press.
- Prinz, J. 2004b. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Prinz, J. March 2006. The emotional basis of moral judgments. *Philosophical Exploration* 9(1): 29–43.
- Prinz, J. (2007) Was William James right about emotions? Paper presented at the *Dutch-Flemish Society for Analytic Philosophy (VAF) Conference*, University of Antwerpen, 26 April 2007
- Ratcliffe, M. 2003. *Not Passion's Slave: Emotions and Choice*. Oxford: Oxford University Press, 2003. In: *Notre Dame Philosophical Reviews*. Retrieved from <http://ndpr.nd.edu/review.cfm?id=1360>. Review of Solomon, R.
- Roeser, S. October 2006. The role of emotions in judging the moral acceptability of risks. *Safety Science* 44(8): 689–700.
- Roeser, S. 2010. Intuitions, emotions and gut feelings in decisions about risks: Towards different interpretation of “neuroethics”. *The Journal of Risk Research* (forthcoming).
- Sartre, J. -P. 1948. *The Emotions: Outline of a Theory*. Translated by B. Frechtman New York: Philosophical Library.
- Slovic, P. 1999. Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis* 19(4): 689–701.
- Slovic, P., M. L., Finucane, E., Peters, and D. G., MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24(2): 311–322.
- Smith, A. January 2005. Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115: 236–271.

- Solomon, R. C. 1980. Emotions and choice. In *Explaining Emotions*. A. Rorty, ed., 251–281, Los Angeles: University of California Press.
- Solomon, R. C. 2003. *Not Passion's Slave: Emotions and Choice*. Oxford: Oxford University Press.
- Solomon, R. C. 2006. *True to Our Feelings: What Our Emotions Are Really Telling*. Oxford: Oxford University Press.
- Tversky, A., and D., Kahneman. September 27, 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185(4157): 1124–1131.
- van Reijen, M. 1995. *Filosoferen over emoties*. Baarn: Nelissen.
- van Reijen, M. 2005. Ik denk dus ik voel: Een stoïcijnse filosofie van emoties. In *Emoties: Van stoïcijnse apatheia tot heftige liefde*. M. van Reijen, ed., Kampen: Klement.
- Young, L., and M., Koenigs. 2007. Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin* 84: 69–79.
- Zhu, J., and P., Thagard. 2002. Emotion and action. *Philosophical Psychology* 15(1): 19–36.

Emotional Reflection About Risks

Sabine Roeser

... we make many claims for the affect heuristic, portraying it as the centerpiece of the experiential mode of thinking, the dominant mode of survival during the evolution of the human species. However, like other heuristics that provide efficient and generally adaptive responses but occasionally lead us astray, reliance on affect can also deceive us. Indeed, if it was always optimal to follow our affective and experiential instincts, there would have been no need for the rational/analytic system of thinking to have evolved and become so prominent in human affairs (Slovic et al. 2002, p. 416).

1 Introduction

Emotions can mislead us in our judgments about risks. They can blur our understanding of quantitative information about risks, but they can also bias us in our judgment of the evaluative aspects of risk. In the literature on risk and emotion, the emphasis is on the former phenomenon. That is why most authors propose that if necessary, risk-emotions should be corrected by rational and scientific methods. However, when it comes to emotional biases of our moral understanding of risks, it is far from obvious that pure rationality will help us out. In this paper I will discuss both kinds of biases. I will argue that not all supposedly emotional biases about the quantitative aspects of risks are really due to emotions, and not all biases are really biases after all. If emotions bias our quantitative understanding of risk, we indeed need proper (accessibly presented) quantitative information. However, concerning the second kind of bias, concerning the moral evaluation of risks, I will argue that we need emotions in order to correct our immoral emotions.

S. Roeser (✉)

Department of Philosophy, Delft University of Technology, Delft, The Netherlands
e-mail: s.roeser@tudelft.nl

2 The Blind Spots of Risk-Emotions

... the affect heuristic enables us to be rational actors in many important situations. But not in all situations. It works beautifully when our experience enables us to anticipate accurately how we will like the consequences of our decisions. It fails miserably when the consequences turn out to be much different in character than we anticipated (Slovic et al. 2002, p. 420).

Apparently emotions are an important guide when it comes to determining our preferences, or when we make value judgments. But Slovic et al. seem to suggest that emotions can be prejudiced and not open for new information. Several authors who write about emotions and risk emphasize the tendency of emotions to blur our vision for certain aspects of risk. As Loewenstein et al. (2001, p. 271) write: “the risk as feeling hypothesis posits that ... emotions often produce behavioral responses that depart from what individuals view as the best course of action”. As Slovic et al. summarize:

Among the factors that appear to influence risky behaviors by acting on feelings rather than cognitions are background mood (e.g., Johnson and Tversky 1983, Isen 1993), the time interval between decisions and their outcomes (Loewenstein 1987), vividness (Hendrickx et al. 1989), and evolutionary preparedness (Loewenstein et al. 2001).

In this section I will discuss the “blind spots” that the various authors have identified. I will examine whether all these blind spots are indeed due to emotions. In so far as they are, I will examine in the remainder of the paper how these blind spots of emotions can be corrected. My claim will be that while some of these blind spots have to be corrected by “rational” or scientific methods, others should be corrected by other emotions.

2.1 *Emotions and Risk Attitudes*

The first bias that I wish to discuss is the general observation made by various scholars that emotions very much determine one’s judgments about risks and benefits:

[P]eople base their judgments of an activity or a technology not only on what they *think* about it but also on how they *feel* about it. If their feelings towards an activity are favorable, they are moved toward judging the risks as low and the benefits as high; if their feelings toward it are unfavorable, they tend to judge the opposite – high risk and low benefit. Under this model, affect comes prior to, and directs, judgments of risk and benefit, much as Zajonc proposed (Slovic et al. 2004, p. 315; italics in original).

Hence, a feeling towards an activity determines somebody’s risk judgment. However, even more general moods to a large degree determine one’s judgments about risks and benefits. This was a finding in a study by Eisenberg, Baron and Seligman (1996) which Loewenstein et al. (2001) report about:

The researchers found that trait anxiety was strongly and positively correlated with risk aversion, whereas depression was related to a preference for options that did not involve taking an action (Lowenstein et al. 2001, p. 273).

I think we can explain these findings as follows: depressive people prefer the status quo because this means that no action needs to be performed, which fits the profile of a depressive person, and, not surprisingly, anxious people are risk averse. Hence, somebody's affective traits determine one's risk attitude. Schwarz (2002) also emphasizes the importance of moods for decision making in general.

I think that moods should indeed be seen as a bias, since moods are not directed towards anything in particular. Hence, they are not responses to a risky activity and yet they determine our attitude towards it. However, feelings that are specifically directed towards a possibly risky activity and determine our risk judgments are not necessarily biased. Our emotions are able to pick out evaluative considerations about risk that by definition cannot be captured by more quantitative approaches towards risk (cf. Roeser 2009). It is by now a common place in the sociological and philosophical literature about risk that risk is not only a quantitative notion but also an evaluative notion (cf. e.g. Fischhoff et al. 1981; Shrader-Frechette 1991; Krimsky and Golding 1992), and that risk attitudes of laypeople comprise a richer understanding of risks (Slovic 2000, cf. Roeser 2007 for a normative-ethical defense of this claim). However, it is quite surprising that in the literature on risk and emotion, these points tend to get forgotten and emotions are mainly discussed in relation to quantitative issues about risks. We will also see that concerning the literature on the other biases that I will discuss in this section.

2.2 Probability Neglect or Availability

The next blind spot that I wish to discuss is what Cass Sunstein calls "probability neglect" and what Paul Slovic calls "availability".

Sunstein (2005) argues that emotions are especially prone to let laypeople neglect probabilities:

Probability neglect is especially large when people focus on the worst possible case or otherwise are subject to strong emotions. When such emotions are at work, people do not give sufficient consideration to the likelihood that the worst case will occur (Sunstein 2005, p. 68).

Slovic et al. understand availability as a heuristic that lets us focus on a risk that is easily imaginable, even though it might not be a very important risk. Slovic et al. (2002, p. 414) argue that imagery is more effective than information about relative frequencies:

Availability may work not only through *ease* of recall or imaginability, but because remembered and imagined images come tagged with affect The highly publicized causes [of death, SR] appear to be more affectively charged, that is, more sensational, and this may account both for their prominence in the media and their relatively overestimated frequencies (Slovic et al. 2002, p. 414).

Slovic et al. say here that “available”, frequently published risks are often more sensational, and thereby more appealing to the imagination and more emotional than risks that get less attention in the media, which clouds our perception of reality. Slovic et al. review various studies that indicate that emotions dominate probabilistic thinking when what is at stake has a strong appeal to emotions, and that the opposite is the case if what is at stake is less affectively loaded:

When the quantities or outcomes to which these probabilities apply are affectively pallid, probabilities carry much more weight in judgments and decisions. Just the opposite occurs when the outcomes have precise and strong affective meanings – variations in probability carry too little weight (Slovic et al. 2002, p. 410).

Emotions can blind us for quantitative considerations. For example, people who suffer from fear of flying are focused on plane crashes, even though these are extremely rare.

2.3 Framing

The third blind spot that I wish to discuss is “framing”. “Framing” refers to the phenomenon that the way (risk-)information is presented to a large degree determines people’s evaluations about that information (Tversky and Kahneman 1974; Slovic 2000; Gigerenzer 2002). This is a phenomenon that holds for both laypeople and experts. Tversky and Kahneman (1974) for example let doctors judge if they would recommend a cancer treatment to a patient. One group of doctors got the information about the effectiveness of the treatment in terms of probability of survival, the other group in terms of probability of death, where the information was statistically equivalent. Representation in terms of probability of survival lead to significantly more positive evaluations of the treatment than representation in terms of probability of death. In this example, “framing” seems to be indeed due to emotions, i.e. positive emotions connected with survival and negative emotions connected with death. However, “framing” is not always due to emotions but can also be caused by other possible sources of irrationality. Gigerenzer (2002) shows that Bayesian representations of probabilities are more confusing (for laypeople and experts) than representations in natural frequencies. This has nothing to do with emotions but with the fact that Bayesian representations require more mathematical insight.

2.4 Manipulation

Another blind spot of risk-emotions that Slovic et al. discuss is manipulation. Manipulation is related to framing but it is broader and presupposes that the sender of the information has the intention to steer the receiver of the information in a certain direction, whereas framing can happen without any such intentions.

According to Slovic et al. (2002), affect can misguide us through manipulation by others. For example, people with attractive names are valued higher, background

music in movies conveys affect and enhances meaning, models in catalogs are smiling to convey positive affect to the products they are selling, food products carry “affective tags” such as “new”, “natural” etc in order to increase the likelihood to be bought. GMOs are called “enhanced” by proponents and “Frankenfood” by opponents (Slovic et al. 2002, pp. 416–417). However, are these really emotions or mere gut feelings? And what is the difference between the two? I will come back to this further on.

2.5 Natural Limitations

Another blind spot are so-called “natural limitations” of our understanding of risks. According to Slovic, the experiential system that also comprises affect is subject to inherent biases:

... the affective system seems designed to sensitize us to small changes in our environment (e.g., the difference between 0 and 1 deaths) at the cost of making us less able to appreciate and respond appropriately to larger changes (e.g., the difference between 570 deaths and 670 deaths). Fetherstonhaugh et al. (1997) referred to this insensitivity as *psychophysical numbing*.

Similar problems arise when the outcomes that we must evaluate change very slowly over time, are remote in time, or are visceral in nature (Slovic et al. 2002, p. 418).

Slovic et al. give the example of nicotine addiction: “a condition that young smokers recognize by name as a consequence of smoking but do not understand experientially until they are caught up in it” (Slovic et al. 2002, p. 418). Slovic explains this as follows: “Utility predicted or expected at the time of decision often differs greatly from the quality and intensity of the hedonic experience that actually occurs” (Slovic et al. 2002, p. 419). However, the example of smoking also indicates the failure of the analytical system¹: apparently, our abstract knowledge is often not very effective in guiding our behavior.

2.6 Proportion Dominance

A last blind spot in our thinking about risks that I wish to discuss and that according to Slovic et al. is due to emotions is proportion (or probability) dominance:

Ratings of a gamble’s attractiveness were determined much more strongly by the probabilities of winning and losing than by the monetary outcomes. [...] We hypothesize that these curious findings can be explained by reference to the notion of affective mapping. According to this view, a probability maps relatively precisely onto the attractiveness scale, because it has an upper and lower bound and people know where a given value falls within that range. In contrast, the mapping of a dollar outcome (e.g., \$9) onto the scale is diffuse,

¹Slovic assumes that there are two mental systems, the affective and the analytical system. This is also what defenders of “Dual Process Theory” argue for. In Roeser (2009) I criticize this approach for being overly simplistic.

reflecting a failure to know whether \$9 is good or bad, attractive or unattractive (Slovic et al. 2004, p. 317).

This is an interesting observation. However, I am not sure what it says about rationality. It seems only reasonable to be agnostic about assessing the value of a certain number if the scale and the upper and lower bounds are unknown. Furthermore, I am not sure in how far this phenomenon really says something about the involvement of affect or emotion. What is the empirical evidence that in the case where bounds are known, evaluations are based on emotions? Maybe the explanation is that Slovic et al. equate ratings of attractiveness with emotional ratings, but whether these are really the same is an open question that should be empirically tested. It is not an analytical claim and it is philosophically controversial whether evaluative judgments are made by reason or emotion or both.

To conclude this section: it is clear that there are many blind spots about risks and probabilities, but they are not all as blind as they seem, and they are not all clearly based on emotions, despite claims to the contrary. Often in debates about bounded rationality, the culprit is by definition “emotion”, without further analysis whether it is indeed emotions that undermine our rationality. Not all spontaneous responses are by definition emotional, yet, this seems to be the hidden assumption (for a critique of this, cf. Roeser 2009). In the next section I will first discuss how the aforementioned authors propose to address the blind spots that have been identified, and I will evaluate in how far these proposals seem justified. In Section 4 I will propose an alternative approach for correcting emotions, namely by emotions themselves.

3 Addressing the Blind Spots

Most authors propose to correct “risk as feeling” by “risk as analysis” (cf. Slovic et al. 2004), by for example scientific information. Sunstein thinks that misguiding emotions should be corrected by cost-benefit analysis:

The role of cost-benefit analysis is straightforward here. Just as the Senate was designed to have a “cooling effect” on the passions of the House of Representatives, so cost-benefit analysis might ensure that policy is driven not by hysteria and alarm but by a full appreciation of the effects of relevant risks and their control. If the hysteria survives an investigation of consequences, then the hysteria is fully rational, and an immediate and intensive regulatory response is entirely appropriate (Sunstein 2002, p. 46).

Sunstein presupposes that cost-benefit analysis is an ultimate arbiter when it comes to evaluations of policies and concomitant emotions. However, cost-benefit analysis has been under severe attack (e.g. Fischhoff et al. 1981; Shrader-Frechette 1991; Slovic 2000, the contributions to Asveld and Roeser 2009). I have argued elsewhere that cost-benefit analysis has to be corrected and completed by ethical intuitions that are also present in risk judgments of laypeople (Roeser 2007) and that these ethical intuitions are based on moral emotions (Roeser 2006). There has to be

a reflective process between technocratic and emotional, explicitly ethical assessments of risks. Hysteria can make us blind, but fear can open our eyes for dangers to which we would otherwise not be sensitive. Slovic et al. (2004, pp. 320, 321) as well argue that a technocratic or analytical approach can benefit from an affect-based approach. Affect can be more suited to convey the meaning that sheer numbers fail to communicate. They mention the examples of literary works and works of art that are better suited than statistics in letting us understand the horrors of the Holocaust and other catastrophes.

Sandman proposes the following solutions: 1. teach people about hazards, 2. make serious hazards outrageous, and 3.:

we have to stop contributing to the outrage of insignificant hazards. As long as government and industry manage low-hazard risks in genuinely outrageous ways – without consulting the community, for example – citizens will continue to overestimate these risks and activists will continue to mobilize against them (Sandman 1989, p. 49).

Hence, Sandman seems to suggest that outrage can actually be created or enhanced by concealing information, and it can be taken away by involving the public. Involving the public creates trust (cf. Slovic 1999; Asveld 2009). This is interesting, since often experts tend to not inform the public about scientific data because they think that the public will not understand them anyway, or because they are afraid of a lawsuit in case their estimates turn out to be wrong, or in case a hazard manifests of which they claimed that it was very unlikely. However, if Sandman and Slovic are right, it is in the own interest of experts to involve the public. Little information creates distrust, which can lead people to opt for a precautionary approach: “better be safe than sorry”. If experts are convinced that a certain technology is worth undertaking, they should share with the public their knowledge about the quantitative risks and benefits and also their ethical concerns. The notion of trust brings me to the following point: namely, that emotions should be corrected by emotions.

4 Correcting Emotion Through Emotion

As said previously, not all the biases the various authors discuss really are instances of affect or emotion. Evaluative responses are not necessarily emotional, and neither are all spontaneous responses. However, this seems to be the hidden assumption in a lot of empirical work on risk and emotion. This assumption is in line with dual process theory (DPT), which serves as a theoretical background for much of the empirical work on risk and emotion. According to DPT, we apprehend reality in two different ways: system 1 is rapid, affective and intuitive, system 2 is slow, analytical and rational (cf. Epstein 1994; Slovic 1996, 2002; Stanovich and West 2002). As I have argued elsewhere (Roeser 2009), this is a much too simplistic conception of the relationship between reason and emotion. Not all spontaneous responses are emotional, and not all emotional responses are spontaneous and a-rational. Even in so far as spontaneous responses are emotional, that does not mean that they cannot

be based on reasons. Some responses that initially involved a process of deliberation can get internalized and evoked spontaneously without reflection in every single instance (cf. Gigerenzer 2007).

Not all emotions are spontaneous, and they are not all unreflected gut reactions. Hence, not all claims that the previously mentioned authors make are strictly speaking about emotions. Spontaneous responses can be characterized as “gut reactions”, but those are not the same as the more cognitive, deliberated emotions that can be the product of lengthy processes of reflection (Roeser 2010). Many contemporary philosophers and psychologists who study emotions defend that emotions can be cognitive and can play a role in reflection and deliberation (cf. e.g., Frijda 1987; de Sousa 1987; Greenspan 1988; Solomon 1993; Stocker 1996; Goldie 2000; Ben Ze’ev 2000; Nussbaum 2001; Roberts 2003).

In any case, to the extent that emotions are involved in biases about risks, the question is how we should examine them and in so far necessary, correct them. In cases where emotions blind us for empirical facts, they should be corrected by scientific methods. However, as said before, the notion of risk is not only a quantitative but also an evaluative notion. I have argued elsewhere that emotions are necessary in order to obtain moral knowledge concerning risks (Roeser 2006, 2009). However, this does not imply that emotions are infallible as a normative guide. Emotions can help us to focus on certain salient aspects, but they can also lead us to overlook other aspects. For example, engineers might be misled by their emotions: their enthusiasm about a product can lead them to overlook certain risks. Policy makers might be tempted to overlook risks because of the desire for economic prosperity for their region that is promised by a certain technology. The public might be ill-informed and hence only focus on risks and overlook certain benefits. They might wrongly estimate the purely quantitative amount of a risk because they perceive it as threatening. All involved parties might be biased, and their emotions might reinforce those biases.

While rationalists would claim that we should correct our emotions by reason, subjectivists would claim that emotions should rule. Instead, a cognitive theory of emotions allows for the idea that emotions themselves have critical potential. Reason and emotion should criticize each other, but emotions should also be used to critically examine other emotions, by trying to understand different perspectives through sympathy and empathy. For example, engineers should try to understand the perspective of the public and vice versa, and those who benefit from a technology should try to understand the perspective of those who are potential victims of the technology. Altruistic emotions can help to conquer egoistic emotions which for example play a role in the NIMBY-problem.

On the position that I defend emotions should themselves play a role in the critical examination of our moral views. Emotions are reflective. Feeling insecure about our moral viewpoint reflects that we have doubts whether we are right. Feeling outrage at a violation of a moral norm such as autonomy might reflect that we are rather confident of that norm. But in the light of thorough disagreement, we might consider reassessing our emotional moral belief, by trying out different points of

view through empathy and sympathy, by putting ourselves in somebody else's shoes and feeling compassion with somebody else. Emotions are not infallible guides to knowledge, but this holds for all our cognitive faculties. Even a rationalist cannot claim that reason always gets it right. In this respect, all epistemologies are in the same boat. However, emotions are often considered to be more notoriously misleading than other mental abilities. I think that this is a mistaken view. To the contrary, purely rational beings without emotions could not make proper moral judgments, especially when it comes to concrete moral judgments in particular situations, as is shown by the famous studies by Antonio Damasio (1994). Emotions are necessary for moral knowledge, but they are no guarantee for success. We need to critically examine our emotions, by exploiting the reflective and critical potential of emotions, which is given through their possibility of shifting points of view and caring for the wellbeing of others.

Furthermore, some of our moral emotions might be more prone to doubt than others. Moral emotions in dilemmatic or complex situations are more fallible, which can be reflected by feeling desperate about whether we made the right judgment or by being torn between two different emotions. Emotions are not infallible, but they *can* lead us to see what is morally right, and they are often better in doing so than our purely rational judgments. If we try to assess whether an emotion is correcting or corrupting our rational moral judgment, we need judgment and emotion as well. We might feel uncomfortable and that we are cheating when giving up a rational judgment based on an emotion, but we might also feel forced to reconsider our initial rational judgment and feel relieved once we have brought our judgment in line with our feeling about a certain case. Whereas the former feeling might point to a corruptive emotion, the latter might point to a corrective emotion.

Michael Lacewing (2005) makes a similar argument, based on ideas from psychoanalysis. He argues that we need to examine our emotions through "emotional self-awareness". According to Lacewing, this involves three things: 1. feeling the emotion, 2. being aware of so doing, and 3. normally, feeling a second-order emotional response to it. He adds a "dispositional fourth": an openness to emotions, which he explains as "a readiness to feel and acknowledge what emotions one has" (Lacewing 2005, 68). Through this process of emotional self-awareness we are able "to detect our anxiety which raises the possibility that our emotional response to the situation is being driven by defense mechanisms" (Lacewing 2005, p. 73). This is important because "[e]motions that are the product of defense mechanisms are not appropriate evaluative responses to the world" (Lacewing 2005, p. 73). A purely rationalist approach runs the danger of a form of intellectualization that "defends against anxiety partly by working with denial, isolation, or repression to simply not *feel* the emotion that arouses anxiety, and partly by using various means of avoiding the emotion's implications and personal significance" (Lacewing 2005, p. 75). As Lacewing emphasizes: "[n]ot feeling any emotion does not mean one's thinking is *undistorted*" (Lacewing 2005, p. 76; italics in original). In other words, rationalizations can us much be distortions as emotions. Lacewing argues that even in cases where emotions are disruptive, it can be important to examine why one feels that

emotion instead of just laying it aside. In such a case the emotional self-awareness can be “detached” but still “engaged” (Lacewing 2005, p. 80).

Let us apply these ideas to emotions about risk. When thinking about the question whether we find a risk morally acceptable or not, we should reflect on our emotions about the risk, but also on our emotional responses to these emotions. If we are afraid of a given technology, can this be sustained by further reflection? Does our fear seem genuine to us? By using emotions such as sympathy and empathy, we can take a more general perspective and try to feel with the position of other people who are possible victims or beneficiaries of that technology. Do we think that overall, this technology is acceptable to society or not? It might be that such emotional reflection reveals that I myself feel upset about a certain technology, although I think that it is a desirable technology for society. This might indicate that I am more driven by egoistic views than by genuine moral concerns about that technology. This would be an example of the NIMBY-problem: I am not against the technology per se, I just don’t want it “in my backyard”. But of course if a technology which is overall desirable (the benefits somehow outweigh risks) but has certain negative side-effects these side-effect will have to affect some people, and it is only fair that everybody will at times be affected by these side-effects. If it is a genuine case of egoism, then higher order emotional reflection can point this out and help us overcome our egoism. This is argued for by the economist Robert Frank according to whom altruistic emotions can solve rational choice problems such as free-riding, i.e. not cooperating and taking advantage of others’ cooperating. Sympathy and fellow-feeling can help overcome “cold-blooded” (i.e. supposedly rational) egoism and promote cooperation (Frank 1988). If we understand the NIMBY-problem as a case of free riding, then we can apply Frank’s insights in order to understand how to solve such problems.

Alternatively, our unease with an overall desirable technology might point out that there can be better ways to deal with the negative side-effects than is initially proposed. In that case, the feeling of unease has to be taken seriously since it points to a morally important consideration. For example, it might be the case that risks and benefits are unfairly distributed, that the risks are involuntarily imposed on some people without giving them a chance to have a say in what is happening, that there are other, less risky and comparatively equally beneficial alternatives or that certain side effects might be unlikely, but that they could be so catastrophic that they are simply unacceptable to those who might be their victims. A test here should be to consider our emotional response if we abstract from the idea that we ourselves are the potential victims to imagining another person being the victim. If we still think that it is unfair, it is apparently not just an egoistic emotion.

This is of course tricky because one of the strategies of emotions such as sympathy is to understand the moral value of the situation of another person by imagining oneself in the role of that person, since that makes it easier to see what might be wrong in that situation. And now I am proposing the opposite procedure – maybe this is asking too much of our imaginative capacities. This concern is also supported by the fact that we tend to care more about the wellbeing of near and dear ones than

of distant others.² On the other hand, this might be a rather limited understanding of moral emotions such as sympathy. Nussbaum (2001) emphasizes that sympathy can broaden our “circle of concern”, for example through reading works of fiction. I have defended elsewhere (Roeser and Willemsen 2004) that the purest form of sympathy is directly directed at the other person, without the need for a detour through our personal perspective.

The corrective potential of emotions should also be used in political decision making about risks. Emotions are generally excluded from political decision making (cf. Hall 2005; Kingston and Ferry 2007 for a critique of this). This also holds concerning political decision making about technological risks (Sunstein 2005 defends this; cf. Kahan 2008 and Kahan and Slovic 2006a; Kahan et al. 2006b for a critique). There the emotions are at most accepted as an unfortunate fact of life (Loewenstein et al. 2001, De Hollander and Hanemaaijer 2003; Wolff 2006 defend such a view). Sunstein criticizes policies that are based on fear of terror (cf. Sunstein 2005). However, the problem with such policies is that they don’t take emotions seriously but use them instrumentally in order to serve a specific political agenda. Such policies respond to people’s gut reactions without critical reflection on emotions. In direct contrast, I think that policy making about risky technologies should do justice to emotions as an invaluable source of ethical insight. Emotions should not be neglected or seen as a “given” that cannot be investigated any further, but they should be a trigger for discussion. Democratic decision making should not just be about counting votes. The arguments, reasons and considerations that are revealed by or lie behind emotional responses to technological risks and benefits have to be taken seriously. Of course the emotional responses of people can differ, but disagreement is nearly always a part of collective decision making, whether or not emotions are included. We should accept the possibly diverging emotions of people and discuss the concerns that lie behind them. Considering diverging emotions is an opportunity to develop more balanced judgments. Our emotions are not infallible, just like other sources of knowledge, emotions can also be mistaken. We should critically assess our emotions, but in doing so, we should take into account other emotions, those of ourselves and of other people.

5 Division of Labor: Scientific Information and Emotions

In the above discussion, I have restricted myself to moral emotions, i.e. to emotions that are involved in moral judgments about risks. However, the blind spots that have been mentioned in Section 2 mainly concerned emotions that distort our access to scientific evidence concerning the descriptive aspects of risk, not the normative aspects. There is need for a division of labor: misguided *moral* emotions should be corrected by the emotional procedures described in the previous section, but emotions that make us blind for *descriptive* facts should be corrected by scientific

²Thanks to Anca Gheaus for pressing me on this point.

evidence. It is important that such evidence is communicated in a way such that people can adjust their emotions to the facts, i.e. in an emotionally accessible way (cf. Buck and Davis, this volume). We saw previously that for example “probability neglect” is a notorious emotional bias in risk perception.

However, it is fallacious to think that in each case where probabilities are low, emotional resistance such as fear is irrational. There can be certain risks that have such catastrophic effects that probabilities become less significant. This is even more the case when there are available alternatives. This might play a role in the fear that many people feel towards nuclear energy. A nuclear meltdown might change large parts of our world for good, even though it is extremely unlikely to happen. And there are many sources of sustainable energy that have no such catastrophic side effects at all, they are even cleaner and less risky than conventional sources of energy (I discuss this in more detail in Roeser 2006). This is a good example of how a technocratic approach may lead to what I would like to call “complexity neglect”. By merely focusing on e.g. annual fatalities we might overlook other morally relevant considerations which can be revealed through emotions such as fear. This example illustrates how risk-emotions can be based on reasonable concerns. These concerns should be taken seriously in debates about the acceptability of technological risks.

Note that various authors who write critically about risk-emotions still emphasize that without emotions, we would be without any guidance (often invoking the work by Damasio on the so-called “somatic marker-hypothesis”):

Emotional reactions guide responses not only at their first occurrence, but also through conditioning and memory at later points in time, serving as somatic markers. Patient populations who lack these markers not only have difficulty making risky decisions, but they also choose in ways that turn their personal and professional lives to shambles. Thus, feelings may be more than just an important input into decision making under uncertainty; they may be necessary and, to a large degree, mediate the connection between cognitive evaluations of risk and risk-related behavior (Lowenstein et al. 2001, p. 274).

Hence, risk-emotions may have blind spots, but without emotions we would be completely blind. Apparently, emotions are an indispensable guide in making decisions about risks, but they are not infallible. Scientific methods with which to measure risks are important corrections to emotions if people tend to ignore scientific evidence because they are ceased by their emotions. Emotions and scientific methods should be in a good balance when thinking about risks: where science can inform us about magnitudes, emotions inform us about moral saliences. Both kinds of information are inevitable if we want to make well-grounded judgments about acceptable risks.³

³Work on this paper was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 275-20-007. A slightly revised Dutch version of this paper has been published as Roeser, S. (2009), “Emotionele reflectie over risico’s in de kennissamenleving”, in Wolter Pieters, Marcus Popkema, Bertien Broekhans, Anne Dijkstra, Kees Boersma & Gerard Alberts (eds.), *Gevoel voor kennis; Jaarboek Kennissamenleving*, Amsterdam: Aksant, 121–138.

References

- Asveld, L. 2009. Trust and criteria for proof of risk: The case of mobile phone technology in The Netherlands. In *The Ethics of Technological Risk*. L. Asveld, and S. (R.) Roeser eds., London: Earthscan.
- Asveld, L., and S., Roeser. (Red.) 2009. *The Ethics of Technological Risk*. London: Earthscan.
- Ben-Ze'ev, A. 2000. *The Subtlety of Emotions*. Cambridge, MA: MIT Press.
- Damasio, A. 1994. *Descartes' Error*. New York: Putnam.
- de Sousa, R. 1987. *The Rationality of Emotions*. Cambridge, MA: MIT-Press.
- Eisenberg, A. E., J., Baron, and M. E. P., Seligman. 1996. Individual differences in risk aversion and anxiety. <http://www.sas.upenn.edu/~baron/>
- Epstein, S. 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist* 49(8): 709–724.
- Fischhoff, B., S., Lichtenstein, P., Slovic, S. L., Derby, and R., Keeney. 1981. *Acceptable Risk*. Cambridge: Cambridge University Press.
- Frank, R. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: W. W. Norton.
- Frijda, N. 1987. *The Emotions*. Cambridge: Cambridge University Press.
- Gigerenzer, G. 2002. *Reckoning with Risk*. London: Penguin.
- Gigerenzer, G. 2007. *Gut Feelings: The Intelligence of the Unconscious*. London: Viking.
- Goldie, P. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.
- Greenspan, P. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification*. New York, London: Routledge.
- Hall, C. 2005. *The Trouble with Passion: Political Theory Beyond the Reign of Reason*. New York: Routledge.
- De Hollander, A. E. M. and Hanemaaijer, A. H. eds. 2003. *Nuchter omgaan met risico's*. Bithoven: RIVM.
- Kahan, D. M. 2008. Two conceptions of emotion in risk regulation. *University of Pennsylvania Law Review* 156: 741–766.
- Kahan, D. M., and P., Slovic (2006a), Cultural Evaluations of Risk: “Values” or “Blunders”? Yale Law School, Public Law Working Paper No. 111
- Kahan, D. M., S., Paul, G., John, and B., Donald. 2006b. Fear of democracy: A cultural evaluation of sunstein on risk. *Harvard Law Review* 119: 1071–1109.
- Kingston, R. and Leonard F. (eds.) 2007. Bringing the passions back. In *The Emotions in Political Philosophy*. British Columbia: University of British Columbia Press.
- Krinsky, S. and Golding, D. eds. 1992. *Social Theories of Risk*. Westport: Praeger Publishers.
- Lacewing, M. 2005. Emotional self-awareness and ethical deliberation. *Ratio* 18: 65–81.
- Loewenstein, G. F., E. U., Weber, C. K., Hsee, and N., Welch. 2001. Risk as feelings. *Psychological Bulletin* 127: 267–286.
- Nussbaum, M. 2001. *Upheavals of Thought*. Cambridge: Cambridge University Press.
- Roberts, R. C. 2003. *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.
- Roeser, S. 2009. The relation between cognition and affect in moral judgments about risk. In *The Ethics of Technological Risks*. L. Asveld, and S. Roeser eds., 182–201, London: Earthscan.
- Roeser, S. 2010. Intuitions, emotions and gut feelings in decisions about risks: Towards a different interpretation of ‘neuroethics’. *The Journal of Risk Research* 13: 175–190.
- Roeser, S. 2007. Ethical intuitions about risks. *Safety Science Monitor* 11: 1–30.
- Roeser, S. 2006. The role of emotions in judging the moral acceptability of risks. *Safety Science* 44: 689–700.
- Roeser, S., and M., Willemsen. January 2004. Compassie en verbeelding. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 96(Nr. 1): 53–65.
- Sandman, P. M. 1989. Hazard versus outrage in the public perception of risk. In *Effective Risk Communication: The Role and Responsibility of Government and Nongovernment*

- Organizations*. V. T. Covello, D. B. McCallum, and M. T. Pavlova eds., 45–49, New York, NY: Plenum Press.
- Shrader-Frechette, K. 1991. *Risk and Rationality*. Berkeley: University of California Press.
- Schwarz, N. 2002. Feelings as information: Moods influence judgments and processing strategies. In *Intuitive Judgment: Heuristics and Biases*. T. Gilovich, D. Griffin, D. Kahnemann, eds., 534–547, Cambridge: Cambridge University Press.
- Sloman, S. A. 2002. Two systems of reasoning. In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, et al. eds., 379–396, Cambridge: Cambridge University Press.
- Sloman, S. A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin* 119: 3–22.
- Slovic, P. 1999. Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis*. 19: 689–701.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis* 24: 311–322.
- Slovic, P., M., Finucane, E., Peters, and D. G., MacGregor. 2002. The affect heuristic. In *Intuitive Judgment: Heuristics and Biases*. T. Gilovich, D. Griffin, and D. Kahnemann, eds., 397–420, Cambridge: Cambridge University Press.
- Slovic, P. 2000. *The Perception of Risk*. London: Earthscan.
- Solomon, R. 1993. *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett.
- Stanovich, K. E., and R. F., West. 2002. Individual differences in reasoning: Implications for the rationality debate? In *Heuristics and Biases: The Psychology of Intuitive Judgment*. T. Gilovich, et al., eds., 421–440, New York: Cambridge University Press.
- Stocker, M. with Elizabeth Hegemann 1996. *Valuing Emotions*. Cambridge: Cambridge University Press.
- Sunstein, C. R. 2005. *Laws of Fear*. Cambridge: Cambridge University Press.
- Sunstein, C. R. 2002. *Risk and Reason. Safety, Law, and the Environment*. Cambridge: Cambridge University Press.
- Tversky, A., and D., Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.
- Wolff, J. 2006. Risk, fear, blame, shame and the regulation of public safety. *Economics and Philosophy* 22: 409–427.

Index

Note: The letters ‘f’ and ‘t’ following the locators refer to figures and tables respectively.

A

- Acceptable risk, judgements
- adequate and inadequate emotions, 183
 - emotional attitude, 181
 - emotion, characteristics, 181–182
 - excited feeling, forms of, 181
 - involvement or ego-preference, 182
 - pro or con attitude, 181–182
 - Spinoza’s theory of emotions, 183
 - temporally extended moods, 181
 - See also* Emotionality and rationality, technological risks acceptance
- Acceptance of technological risks,
- discrepancies (experts and laypeople), 177–180
 - Bayesianism, 178
 - “bounded rationality,” 179
 - context of animal behaviour, “relying on instinct,” 179
 - evidences, 179–180
 - high degree of polarization, 178
 - polarization of Bayesians and non-Bayesians, 180
 - public attention, 177
 - “relying on instinct,” 179
 - stubborn non-acceptance, 177
 - systematic decision-theoretical approaches, 179
 - theory of “qualitative characteristics,” 178
 - See also* Emotionality and rationality, technological risks acceptance
- Access consciousness, 129, 133
- Ackerman, F., 10, 166, 170
- Acousto-optic and photo-acoustic techniques, 140
- Acousto-optic monitoring device for blood, 143–145
- acousto-optic technology, 144–145
 - influence on human life, 145
 - “phantom,” 144
 - photoacoustic mammography, 145
 - “pulse oximeter,” 144
 - See also* Ethical imagination, laboratory deliberations
- Acousto-optic monitoring device, risks in, 140,
- 143, 145, 150–153
 - altered emotions, 152
 - capability to have emotions (Nussbaum), 152
 - “educated imagination,” 150
 - hard/soft impacts, 150–151
 - inequality between lives of light/dark skinned diabetes patients, 151
 - See also* Acousto-optic monitoring device for blood
- Adessi, E., 19
- Adler, J. N., 146
- Advantages/liabilities of emotions, 111–117
- correctness, 112
 - Official Stoic doctrine, 112
 - stage fright, 112
 - criteria for evaluating judgements, 111–117
 - experiential immediacy, 114–115
 - concept of knowledge, 115
 - prosopagnosia, 115
 - justification, 112–114
 - person-centered and virtue-indexed, 114
 - undermined by emotions, 113
 - understanding, 115–117
 - akrasia* (incontinence, weakness of will), 116
 - description of patient, 116
 - potential acquaintance-knowledge, 116

- Advertising and emotion, 66
- Affect, 40–41, 43–44, 53, 63, 169, 199, 237
 affective “fight-or-flight” processes, 65
 ARI model, 63–64
 compassion, mean affect ratings, 53f
 impact on nanotechnology risk perceptions, 169f
 numbers and numbness
 workings of our affective system, 49
 personal affectedness, 197
 syncretic cognition, emotion as affect, 62
 and value of human lives, *see* Affect and value of human lives
- Affect and value of human lives
 normative model
 saving of human life, valuing, 43, 43f
 psychophysical model, 44–47
 airport safety study, 46f
 prospect theory, 45
 psychophysical numbing, 45
 research on psychophysical numbing, importance of, 47
 sensory magnitude (Ψ), 44–45
 studies in life-saving interventions, 46
 valuing saving of human lives, 45f
 “Weber’s law,” 44
See also Psychic numbing and genocide
- Affect-Reason-Involvement (ARI) model, 63–64
 affect/reason continuum (*A/R Continuum*), 63
 A/R scores/ratios, 64
 determined by, 63–64
 suggested relationship, 63f
 emotion and reason, relationship between, 63
- African Union peacekeeping efforts in Darfur, 56
- Ahuja, A., 82
- Ainslie, G., 22
- Air leak in Brown’s Ferry Nuclear Plant, 190
- Aisthêsis* (perception), concept of, 108
- Akerlof, G. A., 169
- Akrasia* (incontinence, weakness of will), 18, 116
- Alcohol and tobacco warnings, 74–77
 cultural and perceptual/ecological factors, 74–75
 anti-drunk driving/antismoking campaigns, 74–75
 International Tobacco Control Four Country Study, 75
 implications, emotion in warnings, 76–77
 advertisements, positive/negative emotions, 76
 individual/situational factors, 75–76
 R.J. Reynolds (company), 75–76
- Alhakami, A. S., 119–120, 160, 164, 166
- Ambiguity of “risk”
 analytic or “second track” processes, 17
 intuitive, “first track” processes, 17
- American National Standards Institute (ANSI), 65, 67
- “Amphiboly,” 134
- Anderson, E., 162
- Andreia* (courage), 111
- ANSI, *see* American National Standards Institute (ANSI)
- Anti-drunk driving and antismoking campaigns, 74–75
- “Antisentimentalism,” 199–202
 “anti-sentimentalist,” Sjöberg’s approach
 aspects in three steps, 200–201
 “appraisal,” 201
 “Dread Factor,” 199
 “emotions” as irrational forces, 201
 enhance public participation/avoid expertocracy, 200
 Scheler, M. in emotion theory, 202
 “objective reality,” 199
 severity of consequences, 199
 “values”/“value dimensions,” 201
- “Antisentimentalism” (new), 199–202
- Aoyagi, T., 140, 144, 146
- Appraisal
 circle of emotional, 22–24
 axiological, 23
 emotional validation, circle of, 23
 Epicurean argument, 24
 epistemic feeling, 23
 hyperbolic discounting, 22
 intuitive and analytic systems, 22
 kinds of appropriateness, 23
 “Law of Apparent Reality,” 22
 “Laws of emotion,” 22
 Peak-End Principle, 22
 practical and epistemic rationality, 22
 pragmatism, philosophy of, 23
 value-belief-means-end nexus, 23
 defined by Prinz, 217–218
- Appropriate emotions, 91, 100, 103, 204, 208
- Appropriateness, kinds of, 23
See also Emotional appraisal, circle of
- APS, *see* Aristotle’s practical syllogism (APS)
- Aristotle, 18–20, 99, 109, 111, 116, 122, 131, 222

- Aristotle's practical syllogism (APS), 18–19
- Arrow, K., 95
- Asian disease problem, 7–9, 98
- 2001: A Space Odyssey*, 129
- Asveld, L., 236–237
- 9/11 attacks, 21
- Attitude
- decision-theoretic description, 103
 - towards risk, 95
 - in terms of the curvature of utility function, 96f
- Attribute substitution
- and prototypical cases, 5–6
 - “availability bias,” domain of risk perception, 5
 - dual-process theories of cognition, 6
 - “moral dumbfounding,” 6
 - system I and system II, 6
- B**
- Bandura, A., 69
- Baron, J., 3–4, 7, 9, 14, 46, 232
- Barrett, L. F., 41
- Bartels, D. M., 46
- Bartneck, C., 128, 133
- Bass, R., 47
- Batson, C. D., 42, 55
- Battito, M. F., 146
- Bayesian calculus (BC), 18
- BC, *see* Bayesian calculus (BC)
- Becker, G., 29
- Beck, U., 195
- Behavioral decision theory, 40
- Beiner, L., 76
- Berman, E. J., 74
- Berndsen, M., 160
- Bernoulli, D., 95
- Bernoulli's law, 95
- Betrayals/betrayal risk, 12–14
 - betrayal aversion, 14
 - “betrayal risk,” 13
 - consequence of violation, 13
- “Beyond-design-basis accident,” 102
- Bickler, P. E., 147–148
- Birnbacher, D., 177, 182, 193, 196, 206–208
- Blackburn, S., 204
- Blalock, G., 21
- Blavatsky, P. R., 98
- Blink: The Power of Thinking Without Thinking*, 121
- Block, N., 129
- Bloom, P., 6
- Bora, A., 17
- Bothma, P. A., 146
- Boyer, P., 31
- Braman, D. K., 166–168, 171, 173
- Bratchenia, A., 144
- Breyer, S. G., 159, 169
- “Brute sensations,” 219
- Buck, R., 61, 63–64, 242
- Burnett, R. C., 46
- Bush (President), 37
- C**
- Cacioppo, J. T., 62
- Cahan, C., 146
- “Capability approach,” 142
- Carmeliet, P., 145
- Carroll, L., 25
- Carruthers, P., 20
- Carter, C. S., 64
- Chaiken, S., 62, 65
 - systematic processing, 62
- Champagne, F., 21
- Chaudhuri, A., 63
- Circle of emotional appraisal, 22–24, 32
 - See also* Emotional appraisal, circle of *Civilization and its Discontents* (Freud), 127
- Clark, M. S., 41
- Clifford, W. K., 22–23
- Clinton (President), 37
- Coalition for International Justice, 54
- Cognitive attribution theorist, 62
- Cognitive priority of emotion to risk perception, 164–165
- Cognitive processing, dual modes of, 61–62
 - central rational or “cold-cognitive” judgment process, 62
 - “central route”/“peripheral route,” 62
 - Chaiken's systematic processing, 62
 - “heuristic processing,” 61
- Cognitivism and James, 220–221
 - Goldie beyond cognitivism and James, 225–226
 - “principle of charity,” 226
 - Prinz's adaptation of James, 224–225
 - causal view, 225
 - Gut Reactions*, 224
 - Jamesian view of emotion, 225
 - perception, 225
 - Roeser's cognitivism, 222–224
 - advantage of rational relations view, 223
 - collective constellation, building, 224
 - commitments, 224

- “emotion blinds judgment,” 222
 - “moral/non-moral emotions,” 224
 - rightness/quality of moral judgment, 223
 - Cohen, G. L., 167–168
 - Coke, J. S., 42
 - “Cold-blooded” egoism, 240
 - “Cold-heart heuristic,” 10
 - Collingridge, D., 30–31, 139
 - “Collingridge dilemma,” 30
 - Compassion, 50–54
 - “compassion fatigue,” 51
 - donating to save statistical/identified lives, 50, 50f
 - “identifiable victim effect” (study), 50–52
 - mean affect ratings, 53f
 - mean donations, 50f
 - model depicting psychic numbing, 54f
 - See also* Psychic numbing and genocide
 - “Complexity neglect,” 242
 - “Conceptual connection,” *see* Rational relations
 - “Concern-based construals,” 107, 109, 121, 124
 - “Confidence crisis,” 207
 - Consequentialism, 7, 185
 - Consequentialist decision theory, 103
 - Conspiracy to Murder: The Rwandan Genocide*, 39
 - “Constellation of moral judgment,” 220
 - “Construal,” 108–110, 112–114, 116, 119, 121, 124
 - Consumer Products Safety Commission (CPSC), 68, 73
 - Contextual knowledge, 143
 - Convention for the Prevention and Punishment of the Crime of Genocide, 55
 - Cook, S., 70
 - Cost-benefit analysis, 9–11
 - “cold-heart heuristic,” 10
 - complementary possibility, 10–11
 - Do not knowingly cause a human death*, 10
 - probability, reframing, 10–11
 - tort law, 10
 - Coté, C. J., 146–147
 - Cox, E. P., 65
 - CPSC, *see* Consumer Products Safety Commission (CPSC)
 - The Critique of Judgement*, 132
 - Cropper, M. L., 8
 - Cultural evaluator theory, 162–164, 164f
 - “judgements of value” (Nussbaum), 163
 - grief/fear/anger, 163
 - social meaning of activity, 162
 - “somatic marker” account, 163
 - Cumulative prospect theory, 97
- D**
- Dake, K., 164
 - Dallaire, Roméo, 40
 - Damasio, A., 116, 163, 170, 202, 239, 242
 - Daniel, J., 3, 42
 - Darfur genocide, 40
 - Dark skin and acousto-optics, 145–150
 - dark skin, 145
 - different ethnic groups, 147
 - oxygen-levels, 147
 - poor performance, dark skin at low oxygen levels, 148
 - skin-colour differences, 146
 - ultrasound, use of, 149
 - See also* Acousto-optic monitoring device for blood
 - D’Arms, J., 204
 - Davidson, D., 18
 - de Angeli, A., 128
 - Decision-theoretic advice, 99
 - Decision-theoretic attempts, St. Petersburg paradox, 95–98
 - attitude towards risk, 95
 - curvature of utility function, 96f
 - concept of expected utility, 95
 - criterion of rational choice, 96
 - cumulative prospect theory, 97
 - dogmatic normativism, 98
 - “framing effect,” 97
 - marginal utility of money (Bernoulli’s law), 95
 - mean-risk method (Weirich), 96
 - Menger’s generalisation, 96
 - Menger’s objection, 95
 - “negative utility,” 96
 - normative adequacy of standard decision theory, 98
 - standard decision theory, 98
 - value function according to Kahnemann and Tversky, 97f
 - See also* Risk assessment as virtue
 - Decision-theoretic standard of rationality, 98, 103
 - Decision theory, 40, 91–95, 98–99, 101–104
 - explications of factual deviations, 94f
 - Decision, two standard models of, 18–21
 - “action” and “decision,” concepts of, 18

- Aristotle's "practical syllogism" (APS)
 advantage over BC, 19
 major failings, 19
- Bayesian calculus (BC), 18
 "degrees of belief," 20
 logical form (Aristotle's discovery), 20
 "motions of animals," 19
 "principle of continence," 18
 process leading to any decision, 32
- "Degrees of belief," 19–21
- De Hollander, A. E. M., 241
- Dennett, D., 133
- Deppa, S. W., 65
- "Design-basis accident," 102
- de Sousa, R., 17–18, 21, 23, 83, 214–216, 238
- DeTurck, M. A., 61, 65
- Deutsch, R., 17
- Dever, G. A., 21
- The Diary of Anne Frank*, 49
- Dickens, C., 122
- Dickens, W. T., 169
- Dillard, A., 44, 47, 51, 53
- Diminishing marginal utility of money
 (Bernoulli's law), 95
- Diving-related injuries and warnings, 67–71
 cultural factors, early pool advertisements,
 68–69
 "group norm," 69
 pool industry, 69
 historical perspective, 67–68
 individual/situational factors, 69–71
 natural exploration, 70
 play, 70
PsychInfo, 70
 "risk-takers," 69
 risk taking, 69
 perceptual/ecological factors, 71
- Dogmatic normativism, 94, 98
- Domestic politics, 38
- Donald, B., 173
- Döring, S., 91, 99, 100, 103
- Douglas, M., 25, 164–165, 170, 196, 197
- Downs, C. W., 65
- DPT, *see* Dual process theory (DPT)
- "Dread Factor," 199
- Dual process theory (DPT), 6, 41, 237
- Dubinsky, Z., 38–39
- Dubner, S. J., 21
- Dutch Bird Protection Agency, 49
- E**
- ECAs, *see* Embodied conversational agents
 (ECAs)
- Eckerman, I., 72
- Eckstein, Z., 29
- "Ecological rationality," 14
- EEA, *see* Environment of evolutionary
 adaptation (EEA)
- Egalitarians, 166–167, 171–172
- Egilman, D., 73
- Egoism, 124, 240
- Eisenberg, A. E., 232
- Eisenberg, N., 42
- Elaboration Likelihood Model (ELM), 62
- ELM, *see* Elaboration Likelihood Model
 (ELM)
- Elster, J., 159, 163
- Embodied conversational agents
 (ECAs), 127
- "Emergency paradigm," 216
- Emissions trading, 11–12
 pollution reduction tool, 12
- Emotion
 advantages/liabilities, 111–117
See also Advantages/liabilities of
 emotions
 advertising and, 66
 as aids and obstacles in risky technologies,
see Emotions as aids and obstacles
 in risky technologies
 altered, 152
 appraisal, *see* Emotional appraisal, circle of
 appropriate, 100
 as bias, irrational weigher theory, 161–162
 as byproduct, rational weigher theory,
 161, 161f
 information processing, effects of, 165–167
 intense negative, 100
 involved in risk perception, *see* Risk
 perception, emotions involved in
 and judgements, *see* Judgements about
 acceptable risk
 laws of, 22
 moral, 64, 77, 83, 124–125, 224, 231, 236,
 239, 241
 and rational relations, 218–220
 in risk regulation, *see* Emotion in risk
 regulation
 and risks, reflection about, *see* Emotional
 reflection about risks
 social and moral, 64
 as syncretic cognition, 62–63
 and systematic reasoning, 167–169
 technological risks acceptance, *see*
 Emotionality and rationality,
 technological risks acceptance

- theory, 201
- in warnings, *see* Warnings
- Emotional appraisal, circle of, 22–24
 - axiological, 23
 - emotional validation, circle of, 23
- Epicurean argument, 24
- epistemic feeling, 23
- hyperbolic discounting, 22
- intuitive and analytic systems, 22
- kinds of appropriateness, 23
- “Law of Apparent Reality,” 22
- “Laws of emotion,” 22
- Peak-End Principle, 22
- practical and epistemic rationality, 22
- pragmatism, philosophy of, 23
- value-belief-means-end* nexus, 23
- Emotionality and rationality, technological
 - risks acceptance
 - Bayesian, apart from, 189–192
 - adequacy of Bayesianism, 191
 - choice with elements of uncertainty, 190
 - fault-tree analysis or simulation, 190
 - discrepancies, experts and laypeople, 177–180
 - Bayesianism, 178
 - “bounded rationality,” 179
 - evidences, 179–180
 - high degree of polarization, 178
 - non-acceptance, 177
 - polarization of Bayesians/non-Bayesians, 180
 - public attention, 177
 - “relying on instinct,” 179
 - systematic decision-theoretical approaches, 179
 - theory of “qualitative characteristics,” 178
 - emotion in judgements on acceptable risk, 180–184
 - adequate (active)/inadequate (passive) emotions, 183
 - “emotional attitude,” 181
 - emotion, characteristics, 181–182
 - episodic forms of excited feeling, 181
 - involvement or ego-preference, 182
 - “pro-” or “con-attitude,” 181–182
 - Spinoza’s theory of emotions, 183
 - temporally extended moods, 181
 - integrating qualitative risk factors into
 - Bayesian scheme, 184–189
 - additional factor of decreased perceived security, 189
 - consequentialism, 185
 - control, 187
 - irreversibility, 187
 - “non-classical” brand of utilitarianism, 185
 - “person-trade-off” approach, 185
 - potential for catastrophe, 188
 - “social amplification,” 185
 - utilitarianism, 184
 - voluntariness, 187–188
- Emotional reflection about risks
 - blind spots
 - addressing, 236–237
 - cost-benefit analysis, 236
 - emotions and risk attitudes, 232–233
 - framing, 234
 - manipulation, 234
 - natural limitations, 235
 - probability neglect or availability, 233–234
 - proportion dominance, 235–236
 - solutions by Sandman, 237
 - correcting emotion through emotion, 237–241
 - democratic decision making, 241
 - “emotional self-awareness,” 239
 - NIMBY-problem, 238
 - reason and emotion, 238
 - division of labor, scientific information and emotions, 241–242
 - “complexity neglect,” 242
 - “probability neglect,” 242
 - “somatic marker-hypothesis,” 242
- Emotion in risk regulation
 - on education of emotions, 171–173
 - cultural identity affirmation and expressive over determination, 172
 - “deliberative risk communication,” 173
 - expertise—scientific and moral, 169–171
 - “deliberative democracy,” 170
 - “populist” regime, 169–171
- Emotion-oriented technologies (EOTs), 128
- Emotions as aids and obstacles in risky technologies
 - arguments about risky technologies, 86
 - “cognitive” theories of emotion, 83
 - ethical and moral beliefs of the person (Haidt’s theory), 85
 - ethical intuitionism, 83
 - Harris and Kass, 85
 - Laws of Fear*, 86
 - “luddite bias,” 86

- “moral foundations theory,” 84
- moral realism, 83
- political liberals, 85
- “theological risk,” 86
- “third ways,” 84
- wisdom in emotion of disgust,
 - bioconservatives, 85
- Engelberg, E., 201
- Environment of evolutionary adaptation (EEA), 26
- EOTs, *see* Emotion-oriented technologies (EOTs)
- Epicurus argument against fear of death, 26
- Epstein, S., 41, 237
- Estes, W. K., 20
- Ethical imagination, laboratory deliberations
 - acousto-optic and dark skin, 145–150
 - See also* Dark skin and acousto-optics
 - acousto-optic monitoring device
 - for blood, 143–145
 - possible risks, 150–152
 - “risk” and “good life,” 141–143
 - “capability approach,” 142
 - contextual knowledge for “educated imagination,” 143
 - DNA tests, 142
 - “hard impacts,” 141
 - limitations, 143
 - “probability,” 141
 - “soft impacts,” 141
 - See also* Risks in acousto-optic monitoring device
- Ethical intuitionism, 83
- Evans, D., 81, 83
- Expected utility, concept of, 95
- F**
- “Faustian” culture of the West, 187
- Fear as construal, 110
- Fear as measure of risk, 27–28
 - common sense hypothesis, 27–28
 - formula expressing expected utility, 27
 - voluntary activities, level of risk, 28
 - “war on drugs,” 27
- Fechner, G. T., 44
- Feckler, M. L., 27
- Feiner, J. A., 147
- Fenske, M. J., 42
- Ferguson, E., 198
- Fetherstonhaugh, D., 45–46
- Finucane, M., 119–121
- “First track” processes, 17, 20–21, 31–32
- Fischer, P. M., 74
- Fischhoff, B., 29, 161–162, 199, 233, 236
- Fisher, E., 140
- Fiske, S. T., 41
- Flynn, J., 124
- Food additives, 163
- Foot and mouth disease, epidemic of, 48
- Foreign Operations Bill, 56
- Forgas, J. P., 41
- “Framing effect,” 8–9, 97, 183
- Frank, R., 240
- Frederick, S., 5–6, 8–9, 42, 162
- “Free discourse” and “communicative action,” 221
- Freud, S., 127–128, 137
- Friedman, M., 95
- Friedrich, J., 46
- Frijda, N., 22, 238
- G**
- Gabrielsen, M. A., 67–68, 71
- Galanter, E., 45
- Gaskell, G., 198, 207
- Genocidal regime of Hussein, S., 56
- Genocide
 - in Darfur, 38
 - facing, 54–56
 - genocide convention, 56
 - psychological obstacles, 54
 - system I and II, 55
 - lessons of, 38–40
 - century of genocide, 39t
 - presidents and genocide, 38–39
 - Tsunamis in South Asia, 39
 - See also* Psychic numbing and genocide
- Geopolitics, 38
- Gershoff, A. D., 13
- Gibbard, A., 194, 204–205
- Gigerenzer, G., 3, 14, 22, 179, 234, 238
- Gilbert, D. T., 38, 55
- Gillroy, J. M., 169
- Gilovich, T., 3
- Givel, M., 75
- Gladwell, M., 121
- Global warming, 38, 163, 166–167, 172
- Glover, J., 37
- Goldie, P., 99–100, 127, 130, 136, 200, 214, 225–226, 238
- Golding, D., 233
- Goucke, R., 146
- Gould, S. J., 5
- Graham, G., 199
- Graham, J., 85
- Greene, J., 6, 24, 213, 222

Greenspan, P., 238
 Gusfield, J. R., 165
 Gustason, W., 95
 "Gut feelings," 179, 215, 235
Gut Reactions, 224
 'Gut reactions,' 198, 200–201, 207, 218–220, 224, 228, 238, 241
 Gyorgi, A. S., 44

H

Hacking, I., 93
 Haidt, J., 6, 10, 42, 55, 84–85
 Hall, C., 241
 Hamilton, D. L., 42, 51–52
 Hammond, D., 75
 Hanemaaijer, A. H., 241
 Hansson, S. O., 141, 193, 206
 "Hard impacts," 141, 150
 Hardin, R., 95
Hard Times, 122
 Hare, R. M., 6
 Harrington, R., 31
 Harsanyi, J. C., 185
 Hastings, G., 76
 "Hate," 41
 Hatzimoyisis, A., 131
 Hegeman, E., 163
 Heinzerling, L., 10, 166, 170
 Helm, B., 103–104
High Tide in Tucson, 49
 Hobijn, B., 27
 Hofstadter, D. R., 25
 Holloway, N., 40
 Hooker, B., 7
 Hsee, C. K., 121
 Hu, J., 128
 Hume, D., 42, 81–82, 84–85, 202–204, 214
 Hurricane Katrina in September 2005, 39
 Hursthouse, R., 128
 Hyperbolic discounting, 22, 26

I

"Ideal speech situation," 221
 Information processing, effects of emotion, 165–167
 authenticity of impulse and risk, 165
 availability effect, 166
 group polarization, 167
 hysteria or mass panic, 166
 individual risk perceptions, responsiveness of, 166
 overconfidence bias, 165
 probability neglect, 165
 thought to reflect biasing effect, 166

Intense negative emotions, 100
 "International symbol," 65–67
 International Tobacco Control Four Country Study, 75
Introduction to Decision Theory, 94
 Inverted St. Petersburg game, 91, 92f, 101–104, 101t
 Inverted St. Petersburg paradox, 100–104, 101t
 "beyond-design-basis accident," 102
 consequentialist decision theory, 103
 decision-theoretic description of our attitudes, 103
 "design-basis accident," 102
 "loss aversion," 101
 methodological individualism, principle of, 102
 real life choices, 101
 "risk-averse"/"risk-seeking," 104
 risk aversion, definition, 103
 "virtuous" reference attitude towards risk and "vicious" limits, 104f
 See also Risk assessment as virtue
 Irrational stimuli, 121–124
 egoism or narcissism, 124
 evolutionary preparedness, 123–124
 past history of fear conditioning, 123
 rhetoric and language, 122–123
 Irrational weigher theory, emotions as bias, 161–162, 162f
 "availability cascades," 161
 emotion-pervaded forms of heuristic reasoning, 162
 "execution of learned rules," 162
 "overconfidence" bias, 161
 "probability neglect," 161
 "status quo bias," 161
 "system 1 reasoning"/"system 2 reasoning," 161
 See also Emotion in risk regulation
 Isen, A., 232
 Ishiguro, H., 129

J

Jackson, M., 40
 Jain, R. K., 145
 James, W., 23, 216–219, 221–222, 224–225
 Jeffrey, R. C., 18, 95
 Jenni, K., 46, 50
 Johnson, E. J., 181
 Johnson, N., 146
 Jolls, C., 161
 Jubran, A., 146, 148
 Judgemental constellation, 220–221

- back to risk, 227–228
 - “perception-judgment” dichotomy, 227
 - redefining the terms, 227–228
 - risk and technology, 227
 - comparison to interpretations, cognitivism and James, 220–221
 - Goldie beyond cognitivism and James, 225–226
 - Prinz’s adaptation of James, 224–225
 - Roeser’s cognitivism, 222–224
 - See also Cognitivism and James
 - “constellation of moral judgment,” 220
 - emotions and rational relations, 218–220
 - “brute sensations,” 219
 - “rational relations” view to, 219
 - “voluntary necessity,” 219
 - emotions as part of, 220–221
 - “constellation of moral judgment,” 220
 - “free discourse” and “communicative action,” 221
 - “ideal speech situation,” 221
 - “rational relation”/“conceptual connection,” 220
 - “free discourse” and “communicative action,” 221
 - “ideal speech situation,” 221
 - mind and body, activity and passivity, 214–218
 - appraisal, defined by Prinz, 217–218
 - de Sousa and Solomon, views of, 216
 - “embodied appraisal” view, 217
 - “emergency paradigm,” 216
 - “gut feelings,” 215
 - James, William on emotions, 217
 - moral judgment, 218
 - “moral sense” or “moral sentiment,” 214
 - “rational relation”/“conceptual connection,” 220
 - “risk as feeling” vs. “risk as analysis,” 213
 - Judgements about acceptable risk, 180–184
 - adequate (active)/inadequate (passive) emotions, 183
 - emotional attitude, 181
 - emotion, characteristics, 181–182
 - episodic forms of excited feeling, 181
 - involvement or ego-preference, 182
 - pro- or con attitude, 181–182
 - Spinoza’s theory of emotions, 183
 - temporally extended moods, 181
 - See also Emotionality and rationality, technological risks acceptance
 - Judgements and emotions
 - accuracy and subtlety of impressions, 110
 - andreia* (courage), 111
 - anxiety insensitive to probabilities, 121–122
 - concept of *aisthêsis* (perception), 108
 - “concern-based construals,” 109
 - “construal,” 108
 - construals different from non-emotional, 110
 - fear as construal, 110
 - irrational stimuli, 121–124
 - egoism or narcissism, 124
 - evolutionary preparedness, 123–124
 - fear conditioning, history of, 123
 - rhetoric and language, 122–123
 - Mueller-Lyer Illusion, 108
 - propositional structure, 109
 - phantasia*, 108
 - phobias, 111
 - praotês* (mildness, gentleness), 111
 - risk/benefit confounding, 119–121
 - sense perceptions, 108, 110
 - socio-political factors, 124
 - Jungermann, H., 177
- K**
- Kahan, D. M., 31, 85, 159–160, 163–164, 166, 168, 171–173, 241
 - Kahneman, D., 3–8, 19, 21–22, 26, 42, 45, 55, 96–98, 101, 161–162, 177, 183, 213, 234
 - Kamm, F., 9
 - Kant, I., 81, 84–87, 131–132, 134–135, 214, 218–219
 - Kaplow, L., 7
 - Kasperson, R. E., 186
 - Kass, L. R., 82, 85
 - Kates, R. W., 190
 - Keeney, R. L., 185
 - Kelleher, J. F., 146
 - Keller, C., 213
 - Kelman, S., 12
 - Kingsolver, B., 49
 - Kline, R., 140
 - Koehler, J. J., 13
 - Koenigs, M., 25, 213
 - Kogut, T., 50–53
 - Kooyman, R. P. H., 153
 - Kopelman, L. M., 141–142
 - Körding, K., 20–21
 - Korsgaard, C., 134
 - Kosfeld, M., 64
 - Kraemer, F., 195–196
 - Krimsky, S., 233

- Kristof, N., 40
 Krohn, W., 195
 Krücken, G., 195
 Krugman, D. M., 74
 Kuran, T., 5, 161, 166
 Kurzweil, R., 29
- L**
- Lacewing, M., 239
 Lacey, M., 54
 Larin, K. V., 144
 Larkin, P., 24
 Laugherty, K. R., 65
 Law of Apparent Reality, 22
 Laws of emotion, 22
 Lazarus, R., 62
 Ledoux, J. E., 41
 Lee, K. H., 147
 Leist, A., 191
 Lengfelder, E., 196
 Lessig, L., 162
 Lessing, D., 33
 Lev, A., 145
 Levi, I., 20
 Levitt, S. D., 21
 “Libertarian paternalism,” 33, 170
 Liénard, P., 31
 Lifton, R. J., 53
 Loewenstein, G., 41, 46, 50, 121, 159,
 161–162, 164–167, 232, 241
 “Loss aversion,” 8, 97, 101
 Lucretius, 23
 Luhmann, N., 195–196
 Luriia, A. R., 25
 Lynch, W. T., 140
- M**
- MacDorman, K., 129
 Mackie, J. L., 203
 MacKinnon, D. P., 75
 MacLean, P. D., 64
 Malek, J., 141–142
 Manes, A., 195
 Manohar, S., 140, 145
 Marketing risk
 alcohol and tobacco warnings, 74–77
 diving-related injuries and warnings, 67–71
 emotion and reason in persuasion/risk
 perception, 61–67
 mattress safety, polyurethane foam/fire
 danger, 72–74
 anti-warnings, 73
 CPSC, 73
 “non-flame-retardant,” 73
 “smoke inhalation,” 73
 Union Carbide pesticide plant disaster,
 Bhopal, 72
 warnings to mattress manufacturers, 73
 Markowitz, H. M., 96
 Martin, B. J., 65
 Martin, R. M., 26, 95
 MacDowell, J., 99, 104, 200, 204
 MacGrath, J. M., 65
 MacIntosh, P. L., 171
 MacLean, A., 64
 MacNaughton, D., 99
 Mean between two “vices,” 99
 Mean-risk method, 96
 Mellers, B., 5
 Melvern, L., 39
 Menger, K., 95–97
 Menger’s generalisation of St. Petersburg
 paradox, 95–96
 Messick, D., 3
 Metacognition, effects of, 28–29
 Bayesian formula, recursive
 fear of fear, increase in, 28
 distortions in perception of risk, 28
 fear of fear, increase in, 28
 Methodological individualism, principle
 of, 102
Metropolis, 134
 Middelkoop, B. J. C., 149
 Milgram, S., 128, 130
 Miller, G. A., 122–123
 Miller, P., 42
 Mill, J. S., 6–7
 Mind and body, activity and passivity, 214–218
 appraisal, defined by Prinz, 217–218
 de Sousa and Solomon, views of, 216
 “embodied appraisal” view, 217
 emergency paradigm, 216
 gut feelings, 215
 James, W. on emotions, 217
 moral judgment, 218
Modern Times, 134
 “Monty Hall problem,” 26
 Moral emotions, 64, 77, 83, 124–125, 224,
 231, 236, 239, 241
 Moral framing and Asian disease
 problem, 7–9
 first component, 8–9
 regulatory interventions, 9
 Moral heuristics and risk
 Asian disease problem, *see* Moral framing
 and Asian disease problem
 heuristics and morality, 6–7

- “corollaries from principle of utility,” 6
 - deontologists, against consequentialism, 7
 - “weak consequentialism,” 7
 - morality and risk regulation, 9–14
 - See also* Risk regulation and morality
 - ordinary heuristics and insistent homunculus, 4–6
 - heuristics and facts, 4–5
 - Moral intuition, definition, 8, 12, 42, 55, 83, 85
 - Morality
 - and heuristics, 6–7
 - “corollaries from the principle of utility,” 6
 - deontologists, against consequentialism, 7
 - “weak consequentialism,” 7
 - and risk regulation, *see* Risk regulation and morality
 - Moral risks of risky technologies
 - access consciousness, 129
 - ambivalence towards technologies, 128–129
 - embodied conversational agents (ECAs), 127
 - instrumental and non-instrumental value, 131–132
 - aesthetic value, 132
 - source of non-instrumental value, 131–132
 - moral reasons for valuing technologies, 133–137
 - “analogues” of humanity/non-human animals (Kant), 134
 - “arbitrary result”/“horrible result” argument (slippery slope argument), 135–136
 - intelligence, 133
 - irritable and short-tempered behaviour, 133
 - personality traits, 134
 - psychological slippery slope, 136
 - phenomenal consciousness, 129
 - rationality of responses to technologies, 130–131
 - paradox of fiction, 130
 - “Moral sense” or “moral sentiment,” 214
 - Mori, M., 128–129
 - Morrison, E. R., 8
 - Mother Teresa, 37
 - Mowrer, O. H., 41
 - Mueller-Lyer illusion, 94, 108–110
 - propositional structure, 109
 - Myers, D. G., 5
 - Myllyläe, R. A., 144
- N**
- Nagel, T., 129
 - Nanotechnology, case of, 30
 - Narcissism, 124
 - Nass, C., 128
 - National Compass for Public health, 149
 - National Swimming Pool Institute (NSPI), 67
 - “Nature bonus,” 145
 - “Negative utility,” 96
 - Neosentimentalism, 195, 204–205
 - Neosentimentalist, 202, 204–206, 208–209
 - Neville, P., 47–48
 - Newcomb’s problem, 26
 - New York Times*, 40, 54–55
 - Nichols, S., 204
 - Night*, 49
 - Nisbett, R., 25, 122
 - Nohre, L., 75
 - Nohrstedt, S. A., 207
 - Nord, E., 185
 - Normative adequacy, 93
 - of decision theory, 93
 - of standard decision theory, 98
 - Normativism, dogmatic, 94, 98
 - Not Passion’s Slave*, 215
 - Nozick, R., 26
 - NSPI, *see* National Swimming Pool Institute (NSPI)
 - Nuclear waste, 29, 163, 199, 201, 221
 - Numbers and numbness, 47–49
 - cost of rescue attempts, 49
 - Dutch Bird Protection Agency, 49
 - foot and mouth disease, epidemic of, 48–49
 - images, 47–48
 - rescue of baby Jessica, 48f
 - September 11 attacks, 48
 - workings of our affective system, 49
 - Nussbaum, M., 15, 83, 142, 152, 159, 163, 171–172, 201–202, 238, 241
- O**
- Objectivity, 17, 222
 - Obsessive compulsive disorder (OCD), 31
 - OCD, *see* Obsessive compulsive disorder (OCD)
 - Official Stoic doctrine, 112
 - Ordinary heuristics and insistent homunculus, 4–6
 - attribute substitution and prototypical cases, 5–6

- “availability bias,” domain of risk perception, 5
- dual-process theories of cognition, 6
- “moral dumbfounding,” 6
- system I and system II, 6
- heuristics and facts, 4–5
- Otway, H., 207–208
- P**
- Pahner, P. D., 207–208
- Panksepp, J., 62, 64
- Paradox of fiction, 130
- Parks, Rosa, 56–57
- Pascal, B., 23
- Paul, S., 17, 37, 47, 119, 168, 197, 213, 233
- Peacocke, C., 103
- Peak-End Principle, 22, 26
- Peng, K., 25
- “Perception-judgment” dichotomy, 227
- Perrow, C., 202
- Persuasion and risk perception, 61–67
 - factors in dangerous behavior
 - cultural/individual/situational/perceptual, 66–67, 72f
 - high/low road to cognition, 61–64
 - ARI model, 63–64
 - dual modes of cognitive processing, 61–62
 - emotion as syncretic cognition, 62–63
 - social and moral emotions, 64
 - implications for persuasion and risk perception, 64–66
 - advertising and emotion, 66
 - effectiveness of warnings, 64–65
 - emotion in warnings, 65–66
 - See also* Affect-Reason-Involvement (ARI) model; Cognitive processing, dual modes of
- Peters, E., 75, 160, 213
- Peters, H., 195
- Petty, R. E., 62
- Phantasia*, 108, 109
- “Phantom,” 144, 146
- Philipson, T. J., 161
- Phobias, 111, 183
- Photoacoustic mammography, 145
- Piaget, J., 70
- Picard, R., 129
- Pictorial international symbols, 66
- Pizarro, D. A., 6
- Plato, 22, 81
- Pogue, B. W., 145
- Pollution reduction tool, 12
- Posner, R. A., 161, 191
- Powell, C., 56
- Power, S., 39–40, 52, 55–56
- Pragmatism, philosophy of, 23
- Praotês* (mildness, gentleness), 111
- Pratt, J. W., 95
- Priest, G., 25
- The Principles of Psychology*, 217
- Prinz, J., 23, 200, 204, 214, 217, 219, 224–225, 228
- “Probability neglect,” 161, 165, 233, 242
- A Problem from Hell: America and the Age of Genocide*, 39
- “Projectivism,” 203
- Prometheus, Greek tale of, 81
- Proxmire, William, 55
- Psychic numbing and genocide
 - affect, analysis, and value of human lives
 - normative model, saving of human life, 43, 43f
 - psychophysical model, 44–47
 - See also* Affect
 - collapse of compassion, 50–54
 - “compassion fatigue,” 51
 - contributions to individuals/group, 52f
 - donating money to save lives, 50, 50f
 - identifiable victim effect (study), 50–52
 - mean affect ratings and mean donations, 53f
 - model depicting psychic numbing, 54f
 - facing genocide, 54–56
 - genocide convention, 56
 - psychological obstacles, 54
 - system I and II, 55
 - Genocide in Darfur, 38
 - lessons from psychological research
 - ‘affect’, 40–42
 - attention, 42
 - behavioral decision theory, 40
 - dual-process theories of thinking, 41
 - feelings motivating people to help others, 42, 42f
 - system I and system 2, 41–42
 - two modes of thinking, comparison of experiential and analytic systems, 41, 41t
 - lessons of genocide, 38–40
 - century of genocide, 39t
 - Presidents and genocide, 38–39
 - tsunami in South Asia, 39
 - numbers and numbness, 47–49
 - cost of rescue attempts, 49

- Dutch Bird Protection Agency, 49
- foot and mouth disease, epidemic of, 48–49
- images, 47–48
- rescue of baby Jessica, 48f
- September 11 attacks, 48
- workings of our affective system, 49
- postscript, 56–57
- Psychic numbing, psychological research
 - ‘affect’, 40–42
 - ‘affect,’ “dual-process theories” of thinking, 41
 - attention, 42
 - behavioral decision theory, 40
 - feelings motivating people to help others, 42, 42f
 - system 1 and system 2, 41–42
 - two modes of thinking, comparison of experiential and analytic systems, 41, 41t
- See also* Psychic numbing and genocide
- Pugmire, David, 217
- “Pulse oximeter,” 144

- Q**
- “Qualitative risk factors” into Bayesian scheme, 184–189
 - additional factor of decreased perceived security, 189
 - consequentialism, 185
 - control, 187
 - irreversibility, 187
 - “non-classical” brand of utilitarianism, 185
 - “person-trade-off” approach, 185
 - potential for catastrophe, 188
 - process of “social amplification,” 185
 - utilitarianism, 184
 - voluntariness, 187–188
- See also* Emotionality and rationality, technological risks acceptance
- Qualitative risk perception, 197–198
 - constructivist turn
 - catastrophic potential of event, 197
 - involuntariness of exposure, 197
 - naturalness of sources, 197
 - perceived controllability, 197
 - personal affectedness, 197
 - positive epistemic emotion of trust, 198
 - “*Risk Perception and GM Food*,” 198
 - “yuck factor,” 198
- Quality of life, 86, 91, 140–141, 143, 150, 185, 207

- R**
- Rabin, M., 95
- Radford, C., 130
- Ramakant, B., 75
- Ramsey, F. P., 18, 20
- Ranachandran, A., 150
- Rational choice, criterion of, 91, 93–96, 94f, 99–100, 103, 178, 240
- The Rationality of Emotion*, 215
- Rational relations, 214, 222–228
 - and emotions, 218–220
 - “brute sensations,” 219
 - “rational relations” view to, 219
 - “voluntary necessity,” 219
- Rational weigher theory, 161, 161f
 - negative emotion experience, 161
 - See also* Emotion in risk regulation
- Rawls, J., 124
- Raymond, J. E., 42
- Reagan, R., 55–56
- Real life choices, 101
- “Recipient oriented” approach, 206–209
- Reeves, B., 128
- Reflective equilibrium, 23–24
- Relative rationality, 24–27
 - context and framing, 27
 - feeling of rightness, 24
 - judgment of, 25
 - rationality, obligatory and optional, 25–26
 - EEA, 26
 - “Monty Hall problem,” 26
 - Newcomb’s problem, 26
 - “Trolley Problem,” illustration, 24
 - VMPC, 25
- Renn, O., 177, 185, 186
- Rescher, N., 180, 191, 206
- Rescue of baby Jessica, 48, 48f
- Resnick, M., 94
- “*Responsibility for Attitudes: Activity and Passivity in Mental Life*,” 218
- Revesz, R., 8
- Richards Jr., J. W., 74
- Rieger, M. O., 98
- Ries, A. L., 147
- Rip, A., 139
- “Risk as feeling” vs. “risk as analysis,” 213
- Risk assessment as virtue
 - appropriate emotions, 100
 - “Asian disease problem,” 98
 - decision-theoretic advice, 99
 - decision-theoretic standard of rationality, 98
 - “infinite” negative value, 92

- intense negative emotions, 100
 mean between two “vices,” 99
 quality of life, 91
 “standard” of rationality, 99
 St. Petersburg paradox, 92–94
 decision-theoretic attempts, *see*
 Decision-theoretic attempts, St.
 Petersburg paradox
 explications of factual deviations, 94f
 Introduction to Decision Theory, 94
 inverted, as a model of risky
 technologies, 100–104, 101t
 inverted St. Petersburg game, 91, 92f
 Müller-Lyer illusion, 94
 normative adequacy of decision
 theory, 93
 Strange Expectations (Hacking), 93
 theory of rational choice, or decision
 theory, 93
 utility maximising strategy, 99
 virtue and emotion in risk assessment,
 98–100
 “Risk-averse/aversion,” 95, 97, 103–104, 178,
 185, 190–192, 233
 Risk/benefit confounding, 119–121
 Risk perception
 cognitive priority of emotion to, 164–165
 emotions
 “confidence crisis,” 207
 “impartial spectator” (Smith, A.), 204
 metaethical implications, 202–206
 neosentimentalist/neosentimentalism,
 204
 new “antisentimentalism,” 199–202
 normative questions, risk and emotional
 damage, 206–208
 qualitative risk perception, 197–198
 “recipient oriented” approach, 207
 sociology of risk: risk objectivists vs.
 risk constructivists, 195–197
 “subjectivism” or “projectivism,” 203
 See also Risk perception, emotions
 involved in
 persuasion and, 61–67
 See also Persuasion and risk perception
 psychometric model of, 197
 qualitative, 197–198
 constructivist turn, 197
 positive epistemic emotion of trust, 198
 “*Risk Perception and GM Food*,” 198
 “yuck factor,” 198
 “Risk Perception and GM Food,” 198
 Risk perception, emotions involved in
 metaethical implications, 202–206
 “impartial spectator” (Smith, A.), 204
 neosentimentalist/neosentimentalism,
 204
 “subjectivism” or “projectivism,” 203
 new “antisentimentalism,” 199–202
 normative questions, risk and emotional
 damage, 206–208
 “confidence crisis,” 207
 “recipient oriented” approach, 207
 qualitative risk perception, 197–198
 sociology of risk
 Chernobyl accident, 196
 “neosentimentalism,” 195
 risk objectivists vs. constructivists, 195
 simplistic objectivism of risk, 196
 sociology of risk, objectivists vs.
 constructivists, 195–197
 Risk regulation and morality, 9–14
 betrayals and betrayal risk, 12–14
 betrayal aversion, 14
 “betrayal risk,” 13
 consequence of violation, 13
 Punish, and do not reward, betrayals of
 trust, 13
 cost-benefit analysis, 9–11
 “cold-heart heuristic,” 10
 complementary possibility, 10–11
 Do not knowingly cause a human death,
 10
 probability, reframing, 10–11
 tort law, 10
 emissions trading, 11–12
 pollution reduction tool, 12
 Risk regulation, emotion in
 empirical evidence, 164–169
 cognitive priority of emotion to risk
 perception, 164–165
 effects of emotion on information
 processing, 165–167
 emotion and systematic reasoning,
 167–169
 normative and prescriptive implications,
 169–173
 on education of emotions, 171–173
 expertise–scientific and moral,
 169–171
 theories of risk perception/conceptions of
 emotion, 160–164
 cultural evaluator theory,
 162–164
 irrational weigher theory, 161–162
 rational weigher theory, 161, 161f

- Risk-seeking, 104
- Risks in acousto-optic monitoring device,
150–152
altered emotions, 152
capability to have emotions (Nussbaum),
152
“educated imagination,” 150
hard/soft impacts, 150–151
inequality between lives of light/dark
skinned diabetes patients, 151
See also Acousto-optic monitoring device
for blood
- Risky technologies
application to, 29–31
case of nanotechnology, 30
“Collingridge dilemma,” 30
“naturalness,” 30
“precautionary principle,” 30
“social construction” of risk, 30
“unknown unknowns,” 30
emotions as aids and obstacles in
arguments about risky technologies, 86
“cognitive” theories of emotion, 83
ethical and moral beliefs of the person
(Haidt’s theory), 85
ethical intuitionism, 83
Harris and Kass, 85
Laws of Fear, 86
“luddite bias,” 86
“moral foundations theory”, 84
moral realism, 83
political liberals, 85
“theological risk,” 86
“third ways,” 84
wisdom in emotion of disgust,
bioconservatives, 85
See also Technologies
- Ritov, I., 14, 50–53
- Roberts, R. C., 32, 107, 118, 121, 238
- “Robo-Rights,” 133
- Roeser, S., 81–84, 125, 197, 202–207, 209,
214, 222, 227, 231–233, 236–238,
241
- Roosevelt (President), 37
- Rosalia, C., 128
- Ross, L., 122
- Rubinstein, Y., 29
- S**
- Sager, E., 27
- Salovey, P., 41
- Sandel, M., 12
- Sandman, P. M., 177, 213, 237
- Savage, L. J., 95
- Save the Children, 50, 53
- Scanlon, T. M., 91, 104
- Scarre, G., 132
- Schabas, W., 55
- Schaber, P., 191
- Scheler, M., 202–203
- Schelling, T. C., 50
- Scherer, K. R., 201
- Schöpfer, G., 195
- Schroeder-Hildebrand, D., 47
- Schroeder, P., 47
- Schütz, H., 198
- Schwartz, S. H., 201
- Schwarz, N., 233
- Selb, J., 140
- Selten, R., 179
- Semi intelligent information filters
(SIIFs), 128
- Sen, A., 142
- Sense perceptions, 108, 110
“*A Sensible Subjectivism*”, 205
- Sfez, B., 145
- Shavell, S., 7
- Sherif, M., 68
- Sherman, J. W., 42, 51–52
- Shrader-Frechette, K., 233, 236
- Sieg, A., 140, 144
- Siegrist, M., 160, 198, 207
- Sigdwick, H., 6–7
- SIIFs, *see* Semi intelligent information filters
(SIIFs)
- Singer, P., 133–134
- Sjöberg, L., 194, 199–202, 209
- Skepticism, 17
- Slater, M., 128
- Slooman, S. A., 237
- Slovic, P., 5, 17, 21, 37, 41, 46, 47, 100,
119–120, 124, 159–161, 164, 166,
170, 177–178, 182, 189, 192,
197, 199, 213–214, 224, 228,
232–237, 241
- Slovic, S., 47
- Small, D. A., 50–51, 53
- Smart, J. J. C., 6
- Smith, A., 204–205, 214, 218
- “Social amplification,” 186
- Social and moral emotions, 64
biologically-based attachment systems, 64
triune theory of brain (MacLean), 64
See also Persuasion and risk perception
- “Social construction” of risk, 30
- Social control of technology, 139

“Social rationality,” 202
 Sociology of risk
 Chernobyl accident, 196
 “neosentimentalism,” 195
 risk objectivists vs. constructivists,
 195–197
 simplistic objectivism of risk, 196
 Socrates, 22–23
 “Soft impacts,” 141–142, 150–151
 Solomon, R., 200–202, 214–216, 225, 238
 “Somatic marker-hypothesis,” 242
 Song, J., 49
 Spivey, M., 67–68, 71
 Standard decision theory, 98
 Standard of rationality, 98–99, 103, 179, 207
 Stanovich, K. E., 17, 41, 55, 237
 Starr, C., 28, 187, 189, 197
 Steinfath, H., 204
 Stevens, S. S., 44
 Stewart, M., 40
 Stocker, M., 132, 163, 217, 238
 St. Petersburg game, 92, 92t
 St. Petersburg paradox, 92–94
 explications of factual deviations, 94f
 Introduction to Decision Theory, 94
 inverted St. Petersburg game, 91, 92f
 Müller-Lyer illusion, 94
 normative adequacy of decision theory, 93
 Strange Expectations (Hacking), 93
 theory of rational choice, or decision
 theory, 93
 See also Risk assessment as virtue
 Strack, F., 17
Strange Expectations, 93
 “Subjectivism,” 124, 203–205
 Sunstein, C., 3, 5, 9, 11–12, 31, 33, 86, 159,
 161–162, 165–166, 169–170, 172,
 233, 236, 241
 Susskind, J., 42, 51–52
 Syncretic cognition, emotion as, 62–63
 affect, definition, 62
 central memory networks, LeDoux, 62
 Lazarus, R. (cognitive attribution
 theorist), 62
 memory and processing systems in
 brain, 62
 See also Persuasion and risk perception
 Systematic reasoning and emotion, 167–169
 hypothesized interactions of information
 and emotion, 168f
 impact of affect on nanotechnology risk
 perceptions, 169f
 risks of nanotechnology, 168

T

Tappolet, C., 100, 103
 Taubert, N. C., 196
 Taylor, T. M., 76
 Technologies
 ambivalence in behaviour towards,
 128–129
 application to risky, 29–31
 case of nanotechnology, 30
 “Collingridge dilemma,” 30
 “naturalness,” 30
 “precautionary principle,” 30
 “social construction” of risk, 30
 “unknown unknowns,” 30
 emotions as aids and obstacles in
 arguments about risky technologies, 86
 “cognitive” theories of emotion, 83
 ethical and moral beliefs of the person
 (Haidt’s theory), 85
 ethical intuitionism, 83
 Laws of Fear, 86
 “luddite bias,” 86
 “moral foundations theory”, 84
 moral realism, 83
 political liberals, 85
 “theological risk,” 86
 “third ways,” 84
 wisdom in emotion of disgust,
 bioconservatives, 85
 instrumental and non-instrumental value
 of, 131–132
 aesthetic value, 132
 source of non-instrumental value,
 131–132
 moral reasons for valuing, 133–137
 “analogues” of humanity/non-human
 animals (Kant), 134
 “arbitrary result”/“horrible result”
 argument (slippery slope argument),
 135–136
 intelligence, 133
 irritable and short-tempered behaviour,
 133
 personality traits, 134
 psychological slippery slope, 136
 moral risks of risky, *see* Moral risks of
 risky technologies
 rationality of our responses
 to, 130–131
 paradox of fiction, 130
 Tetlock, P., 10
 Teuber, A., 141
 Thaler, R. H., 33

- Theory of rational choice, 93, 95, 100
See also Decision theory
- Tobin, M. J., 146, 148
- Todd, P., 3
- Tomkins, S. S., 41
- Tort law, 10
- Townsend, E., 198, 207
- Trapp, R. W., 185
- Travesty of the Golden Rule, 192
- “Treasure,” 41
- Tromberg, B. J., 145
- Tsiddon, D., 29
- Tsunami (December 2004), 37
- Tucker, D.M., 62–63
- Tversky, A., 3–4, 7–8, 21, 45, 96–98, 101, 177, 181, 183, 186, 213, 232, 234
- Tyndall Report, 40
- U**
- Ubel, P. A., 46
- UK Office of Science and Innovation’s Horizon Scanning Centre, 133
- University of Twente, 140, 144, 147–148
- U.N. peacekeeping mission in Rwanda, 40
- Upheavals of thought*, 152
- Utility maximising strategy, 99
- V**
- Value-belief-means-end* nexus, 23
- Value function
 according to Kahnemann and Tversky, 97f
 “virtuous” reference attitude and “vicious”
 limits of admissible corridor, 104f
- Value of human lives and affect
 normative model
 saving of human life, valuing, 43, 43f
 psychophysical model, 44–47
 airport safety study, 46f
 prospect theory, 45
 psychophysical numbing, 45
 research on psychophysical numbing,
 importance of, 47
 sensory magnitude (Ψ), 44–45
 studies in life-saving interventions, 46
 valuing saving of human lives, 45f
 “Weber’s law,” 44
See also Psychic numbing and genocide
- van der Burg, S., 139–140
- van der Pligt, J., 160
- Van Lear, C. A., 63
- Västfjäll, D., 53
- Ventromedial prefrontal cortex (VMPC), 25
- Virtue and emotion in risk assessment, 98–100
- “Virtue is knowledge” (Plato), 22
- Virtuous risk assessment, *see* Risk assessment
 as virtue
- Viscusi, W. K., 10, 161, 169
- VMPC, *see* Ventromedial prefrontal cortex
 (VMPC)
- “Voluntary necessity,” 219
- Vuilleumier, P., 42
- W**
- Walters, R., 69
- Wang, L. V., 98, 140, 145
- Wang, M., 98
- Warnings
 alcohol and tobacco, 74–77
 anti-drunk driving/antismoking
 campaigns, 74–75
 implications, emotion in warnings,
 76–77
 individual/situational factors, 75–76
 International Tobacco Control Four
 Country Study, 75
See also Alcohol and tobacco warnings
- diving-related injuries and, 67–71
 cultural factors, early pool
 advertisements, 68–69
 historical perspective, 67–68
 individual/situational factors, 69–71
 perceptual/ecological factors, 71
See also Diving-related injuries and
 warnings
- emotion in, 65–66
 affective “fight-or-flight” processes, 65
 definition, 64
 effective, warnings, 66
 pictorial international symbols, 66
- “War on drugs,” 27
- “Weak consequentialism,” 7
- Weber, E. H., 44
- Weber, E. U., 121
- Weijers, R. N. M., 149
- Weirich, P., 95–97, 99
- West, R. F., 41, 55
- White, R. W., 70
- Whitlow, J. W. J., 20
- Wiedemann, P. M., 198
- Wiesel, E., 49
- Wiggins, D., 200, 204–205
- Wildavski, A., 197
- Willemsen, M., 241
- Williams, B., 135–136
- Wilson, T. D., 21
- Wise Choices, Apt Feelings*, 205
- Wogalter, M. S., 65

Wolff, J., 241
Wolpert, D. M., 20
Wood, W. J., 121, 143
World Health Organization, 74
World War II, 39, 55, 143

X

Xu, M., 140

Y

Yaak Valley in Montana, 47

Young, L., 213
“Yuck-factor,” 198

Z

Zajonc, R. B., 28
Zhao, Z., 144
Zuckerman, A., 65
Zuckerman, M., 69
Zwart, S. D., 140
Zwick, M., 177, 185