Emotions and Risky Technologies

Author(s) Roeser, Sabine; Roeser, Sabine

Imprint Springer Netherlands, 2010

ISBN 9789048186471, 9789048186464

Permalink https://books.scholarsportal.info/uri/ebooks/ebooks2/springer/2011-02-17/2/9789048186471

Pages 17 to 35

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

[1] What 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.

S. Roeser (ed.), *Emotions and Risky Technologies*, The International Library of Ethics, Law and Technology 5, DOI 10.1007/978-90-481-8647-1_2,

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 (Addessi 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 uninspected 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 Körding'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 risky[2]) 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.

---

[2]Depending 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).

[3]Some 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 dammed). 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).

---

[4]Pascal 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–>q)* should compel us to believe *q*, or to reject *either p or (p–>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 judgements 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

[5] A 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 (Luriia 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]

---

[6] This 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).

[7] You 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 is 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.

[8]This 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:

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

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.

[9]Economically 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 unpredicted (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 "firewall" 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

[10] As 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

Addessi, 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 Suplement*, 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.

Luriia, 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, Indianapol: 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.