

Rethinking Rationality

To make its research readily available to a broad audience, the Institute for Philosophy and Public Policy publishes a quarterly newsletter: *QQ—Report from the Institute for Philosophy and Public Policy*. Named after the abbreviation for “questions,” *QQ* summarizes and supplements Institute books and working papers and features other selected work on public policy questions. Articles in *QQ* are intended to advance philosophically informed debate on current policy choices; the views presented are not necessarily those of the Institute or its sponsors.

* * *

In this issue:

Recent research on the scope and limits of rationality triggers worries about the wisdom of our public policies governing risk p. 1

Does the scientific basis for human genetic engineering justify faith that it will not create more problems than it can solve? p. 6

Does nuclear deterrence work? For an answer we need to take a closer look at the nature of deterrence itself p. 9

How good a person do I have to be? (How good a person do I *want* to be?) p. 12

The first volume in a new book series is announced p. 15

Just when you were getting used to the idea that alfalfa sprouts cause cancer and exercise causes infertility comes word about the possible threat posed by radon in your home. Every night of late local news stations in Washington, D.C., advertise their evening report by a teaser promising more information on the threat posed by this invisible form of indoor air pollution. The estimated chance of death by radon, even on the most pessimistic accounts, is far smaller than the chance of dying in a car accident; yet few news broadcasts try to attract viewers by headlines promising defensive driving tips. Everyone is afraid of getting AIDS; far fewer dread diabetes, a much more prolific killer. One survey shows that over 90 percent of us think we are better than average drivers.

It is well known that people's worries about various risks correlate poorly with the actual dangers they pose. Our judgments are seriously flawed, and it seems that, at the very least, our attitudes about risk are often inconsistent, if not perverse.

Of course, those charged with inconsistency in such matters are free to respond, “So I’m inconsistent. Big deal!” Many would join with Emerson in holding that “consistency is the hobgoblin of little minds,” or with William Allen White, that “consistency is a paste jewel that only cheap men cherish.” We may have our own reasons to pick and choose our own fears, whether or not our choices line up with somebody else’s dispassionate assessment of what is truly fearsome.

Yet consistency in some form defines what it is to be rational, and we are less complacent about the charge

of irrationality. Certain principles of consistency in our choices—so obvious, on first reflection, that they hardly need stating—have been taken to be constitutive of what rationality is. If risk is what you care about, and automobiles are riskier than radon in the home, isn't it irrational not to adjust your concerns about the two accordingly?

People talk about risk, and experts talk about risk, but the two groups may be talking past each other, for "risk" often seems just a convenient label to slap on a broad family of other concerns.

If it turns out that people are indeed irrational in their attitudes toward risk, this has troubling implications for many of our public policies governing risk. In a democracy public attitudes are the cornerstone for policies; they are embodied in the laws we pass and reflected in regulations that give those laws substance. If public opinion on regulatory issues is shaped by frivolous or confused considerations, this bodes ill for our prospects of establishing rational public policies.

Perhaps, however, apparent inconsistencies in our attitudes about risks can be given some other explanation. Perhaps people focus concern on certain risks because they are responding to factors that researchers are simply failing to measure. Or perhaps the standard model of rationality cannot accommodate the complexities of human reasoning. How rational or irrational are we? And how much does this matter?

Who Cares About Risk Anyway?

A first, sympathetic explanation of why people worry disproportionately about certain risks—that is to say, out of proportion to the riskiness of the risk—is that the category of risk hardly exhausts the full range of what most of us care about in our daily lives. People talk about risk, and experts talk about risk, but the two groups may be talking past each other, for "risk" often seems just a convenient label to slap on a broad family of other concerns.

Risk analysts compute the chance of dying from a given activity, but most people care not only about their chance of dying but about what life is like while they are living it. Trade-offs between quality and quantity of life are made all the time in personal decisions about health and safety: Mark Twain, told that he could add five years to his life by giving up smoking and drinking, reportedly quipped that five years without smoking and drinking weren't worth living. Similar trade-offs are relevant in the policy arena as well.

Thus people may prefer one technology to another for reasons other than the actual risks to life and health associated with it. They tend to care about the form

of social organization it encourages (solar power lends itself to decentralization, while nuclear power is by its nature highly centralized); about the control they feel they have over its risks (one, perhaps specious, reason for worrying less about driving than flying); about whether the risks are assumed voluntarily or imposed by others, whether these occur now or later, affect many or few, and so forth.

These factors help to explain some of the results the experts find so puzzling. If risk analysts are preoccupied exclusively with risk, while our concerns are more diverse and wide ranging, it is not surprising that their research should deliver a verdict of irrationality. But here the fault lies not in how human beings think about risk, but in an overly narrow and restrictive focus of measurement.

Other findings are less easily explained, however, by pointing to a richness in our values overlooked by risk analysts. They suggest that people are often driven by indefensible cognitive processes; at least some of the time we simply process information in crazy ways.

Failing Grades in Probability

Measurements of risk have two components: an assessment of the probability of some outcome and an assessment of whether that outcome would be good or bad, and how good or bad it would be. Thus opportunities arise for people to make two different kinds of mistakes.

First, we can make erroneous judgments about probabilities. Many of the laws of probability have a counter-intuitive flavor, and temptations to fallacy are common. Most of us have to struggle to resist the so-called gambler's fallacy, the wistful belief that after enough successive losses the odds have to favor a win. Only reluctantly do we abandon our gut feeling that after five coins in a row have come up heads, surely the next coin will come up tails, despite the mathematician's insistence that the odds on any fair coin toss remain fifty/fifty.

Most of us, who can hardly balance our checkbooks without a pocket calculator, wisely eschew complicated calculations of probabilities in favor of simple rules of thumb, or heuristics. . . .

The gambler's fallacy is an obvious and familiar example of a cognitive failure. More interesting is research on how poorly people do at probabilistic reasoning even when they appear to be doing it quite well.

Most of us, who can hardly balance our checkbooks without a pocket calculator, wisely eschew complicated calculations of probabilities in favor of simple rules of

thumb, or heuristics, that help us assess probabilities in a rough and ready way. The pioneering psychologists Daniel Kahneman and Amos Tversky have studied many of these heuristics. One is "salience," or the ease with which we can call examples of a certain kind of occurrence to mind. Salience will be a generally reliable guide to likely patterns in the world around us: we can think of fewer redheads than brunettes in our acquaintance because brunettes indeed outnumber redheads in the population. But such heuristics can lead us to make mistakes, as when the salience of some risk is reinforced by undue media attention (another possible reason why airplane crashes are more feared than automobile accidents). On balance, however, the heuristics Kahneman and Tversky identify, unlike the gambler's fallacy, are promising strategies for coping efficiently with the uncertainties we face. It is not surprising that even the best heuristics will not always give the same results one would get if one took the trouble to figure in the specifics of the case, but it is surely reasonable to sacrifice some accuracy for convenience.

What is surprising, however, is the extent to which these heuristics dominate even obviously relevant countervailing information. Tversky and Kahneman have shown that when heuristics are triggered, people let themselves ignore completely information that may be far more important but less salient. In one study, Tversky and Kahneman asked people to judge whether an individual selected from some population was more likely to be an engineer or a lawyer, where the population consisted of 70 percent engineers and 30 percent lawyers. Such "base rate data" about the background population is crucial to assessing probabilities correctly. They found that in the absence of any description of the individual's characteristics and traits, people's judgments were based on what they knew about the percentage of lawyers and engineers in the population, but "prior probabilities were effectively ignored when a description was introduced, even when this description was totally uninformative." When asked to decide

if some undescribed Tom was more likely to be a lawyer or an engineer, subjects guessed he was 70 percent likely to be an engineer, but when they were asked about Dick, who was described simply as a "30 year old man, married with no children," these subjects opted for a fifty/fifty probability.

The conclusion to draw from such research seems fairly straightforward. We not only rely on heuristics but are dominated by them, even when we shouldn't be. When it comes to judging probabilities people are, with some frequency, simply wrong.

Knowing What We Value

A more controversial and troubling realm of error lies in the way that people value outcomes. We may expect people to be poor at assessing probabilities, but we want to give them credit for at least knowing their own minds when it comes to assigning values to the outcomes of their choices. They can confidently judge which of two alternatives is the more attractive.

Or can they? Another striking finding of Kahneman and Tversky is that people's value judgments are notoriously influenced by the way in which various choices are framed. Every retailer knows that customers consider \$99 a far more attractive price than one a mere dollar higher, while the difference between, say, \$97 and \$98 makes no difference at all. Similarly, customers tend to be delighted by a discount for cash payment but bristle at a surcharge for using credit cards—although identical policies can usually be described either way.

What Tversky and Kahneman have done is to describe the systematic nature of how preferences are affected by the way a problem is framed. They have scientifically grounded the suspicion that people prefer the "half full" cup over the "half empty" cup, showing the extent to which framing does indeed affect judgment, even on some important policy issues.

Kahneman and Tversky ask us to suppose that the United States is preparing for the outbreak of an

A cab was involved in a hit-and-run accident. Two cab companies serve the city: the Green, which operates 85 percent of the cabs, and the Blue, which operates the remaining 15 percent. A witness identifies the hit-and-run cab as Blue. Tests show that under circumstances similar to those on the night of the accident, witnesses correctly distinguish Blue cabs from Green cabs 80 percent of the time and misidentify them the other 20 percent. What's the probability that the cab involved in the accident was Blue, as the witness stated?

Most subjects conclude that it is 80 percent likely that the cab will be Blue. If there were 85 Green cabs and 15 Blue ones in the city, however, a witness with an 80 percent accuracy rate would incorrectly identify 17 Green cabs as Blue (20 percent of 85 = 17), and he would correctly identify 12 of the Blue cabs (80 percent of 15 = 12). He would thus identify 29 cabs as Blue and be correct in only 12 cases, an error rate of almost 60 percent. The base rate—the preponderance of Green—makes the odds 60 to 40 that he has misidentified a Green cab rather than correctly identified a Blue one.

(Example from Tversky and Kahneman)

Linda is 31, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with discrimination and other social issues and participated in anti-nuclear demonstrations. Which statement is more likely:

- a. Linda is a bank teller.**
- b. Linda is a bank teller and active in the feminist movement.**

87 percent of respondents ranked (b) as more likely than (a), but it is a law of probability that a conjunction cannot be more likely than one of its constituents.

(Example from Tversky and Kahneman)

unusual Asian disease, which is expected to kill 600 people, unless action is taken. Two alternative programs to combat the disease have been proposed. If program A is adopted, 200 people will be saved. If program B is adopted, there is a $\frac{1}{3}$ probability that 600 people will be saved and a $\frac{2}{3}$ probability that no people will be saved. When the alternatives were posed in these terms in a test survey, 72 percent of respondents opted for Program A, only 28 percent for program B. A second group was given the same options, but described in this way: If program A is adopted, 400 people will die; if program B is adopted, there is a $\frac{1}{3}$ probability that nobody will die, and a $\frac{2}{3}$ probability that 600 people will die. This time only 22 percent opted for the first program, while 78 percent opted for the second—a clear preference reversal. The framing of the question proved decisive in eliciting a response. When program A was seen as involving a gain of 200 lives it was rated far more favorably than when it was seen as involving a loss of 400.

This research, moreover, reveals not only the extent to which framing affects what is perceived as a loss or a gain, but a deep and abiding asymmetry in how losses and gains themselves are regarded. In study after study, people show themselves more concerned to avoid a loss than to receive an equivalent gain. In one experiment, for example, researchers Jack Knetsch and J.A. Sinden gave half their subjects tickets to a lottery and the other half \$3.00. When the first group was given an opportunity to sell their tickets for \$3.00, 82 percent kept them, but when the second group was allowed to buy lottery tickets for their \$3.00, only 38 percent wanted the tickets. Human beings seem strongly disposed, for whatever reason, to defend their own personal status quo, to hang on to what they've got. Tversky and Kahneman call this tendency "loss aversion."

Framing effects and loss aversion together explain an important class of preference reversals. People value gains and losses differently, and the framing of a problem determines a "reference point" from which outcomes are viewed as gains or losses. In the Asian

disease example, the description of the problem determines whether people see the alternatives as lives saved or lives lost, which in turn determines preferences among the alternatives.

Is loss aversion irrational? On one popular model of economic rationality it is. According to this conception, what matters is the bottom line, where you end up, whether you got there by avoiding a loss or forgoing a gain. But for others, the choice process itself may legitimately matter to us, as well as its ultimate outcome. Douglas MacLean, director of the Institute for Philosophy and Public Policy's project on risk and rationality, argues that "we have a common set of reactive attitudes, like regret and reproach, which are often provoked by the decisions or choices we make and not just by the choiceless outcomes that result from our decisions. The very same outcome might provoke delight in one context—some small gain you were lucky to realize—but regret in another, for example if you could have done far better by choosing differently." It is not always reasonable to be affected by such attitudes, of course, but given their persistence and pervasiveness, we may do well to factor them into our decisions. For example, MacLean suggests, "if your acquaintances, your boss, or your constituents are more likely to hold you responsible for the losses that result from your risky decisions than for the lost opportunities that result from choosing to avoid risks, then choosing to avoid a loss rather than maximize the expected outcome would seem to be an eminently prudent strategy." An adequate conception of rational choice may have to assess the value of decision processes themselves and our reasonable reactive attitudes and aversions. This would involve a more complex understanding of alternatives than one that exclusively focuses on outcomes in the measurement of risk.

We might accept loss aversion, but what about our

Imagine you have operable lung cancer and must choose between two treatments: surgery and radiation therapy. Of 100 people having surgery, 10 die during the operation, 32 are dead after one year, and 66 after five years. Of 100 people having radiation therapy, none die during treatment, 23 are dead after one year, and 78 after five years. Which treatment would you prefer?

When this question was posed to a group of physicians, framed in terms of mortality rates, 50 percent of respondents favored radiation therapy. But when the same information was presented in terms of survival rates, radiation was preferred by only 16 percent. Surgery is somewhat riskier at the outset, but has better odds of long-term survival.

(Example from Kahneman and Tversky)

susceptibility to framing effects? Surely preference reversals in cases like the Asian disease example are irrational. After all, the two programs of disease control remain the same, however described. Nevertheless, the issue of framing effects might be more complicated. MacLean acknowledges that "if two different descriptions describe the same choices, rationality requires that we make the same decision in each case. Stated in this way, the principle seems hardly exceptionable." But "the problem is that there may be no clear way to determine what counts as a redescription of the same prospect, rather than descriptions of different prospects." Is the situation of someone who gained some money and lost it gambling the same or different as that of someone who never had any money to begin with? While framing can determine the reference point from which different possible outcomes are viewed as gains or losses, there may be no general way to determine what the correct reference point should be. One's current position is not always an appropriate benchmark. A raise in salary, for example, might trigger delight by comparison to one's actual salary, but disappointment compared to one's legitimate expectations.

Much of the moral fabric of our lives depends on finding the appropriate descriptions of objects and events that, from some detached perspective, could equally well be described in some other way. A person's arm can be seen as a human limb or as meat and bones. Assessments of guilt and responsibility, in law and in morality, often turn on determining what description of an event is true or most appropriate. And our culture is currently divided most deeply on whether a human fetus is a person or merely a human organism.

Finally, some recent research in this area suggests further complexities, that people's value judgments apparently depend not only on how the outcomes of a choice are framed, but also on the framing of the choice itself, the process by which the judgments of gain or loss themselves are elicited. Paul Slovic, of Decision Research, Inc., has shown that we will get a different picture of the value people place on, say, tea or coffee, according to whether we ask them whether they *prefer* tea to coffee, whether they *judge* tea to be *better* than coffee, or whether they would be *willing to pay* more for tea than for coffee. This research also reveals some remarkable preference reversals. Some of Slovic's examples focus on a choice between two bets with roughly equal expected payoffs: one has a higher probability of winning, while the other involves a smaller chance of a larger gain. People prefer to take the higher probability bet, but they are willing to pay more for the chance to take the other. Again, according to one popular conception of rationality, values, preferences, and choices should reflect the same consistent ordering, so these reversals would be irrational. But some of Slovic's results might be taken instead to show how values can be expressed in importantly different ways. We might value some things deeply, for example, but think it inappropriate or wrong to pay for them at all.

What Should We Conclude?

The conclusions from this research are mixed. We find some striking examples of human fallibility and some clear limits to human rationality. But other cases show plain folks exhibiting reasoning that may not square with expert assessments of risk but nonetheless does not seem obviously faulty. On the one hand, people do poorly at judging probabilities and are persistently led astray by framing effects. On the other hand, their reasoning often shows both good sense and a sensitive appreciation of the complex nature of difficult choices.

In any case, it is doubtful that people will ever change to become more fully rational. Tversky and Kahneman's research findings, confirmed in countless studies over many years, show that people's patterns of choice are amazingly tenacious and persist at all levels of education and technical sophistication. In a country where most adults would be hard pressed to find a least common denominator between two fractions, education in the laws of probability and in the subtleties of risk analysis is going to produce limited results. We are going to have to deal with people as they are.

On the one hand, people do poorly at judging probabilities and are persistently led astray by framing effects. On the other hand, their reasoning often shows both good sense and a sensitive appreciation of the complex nature of difficult choices.

Nor does increased reliance on expert risk analysis seem to provide a better foundation for public policy. Our irrationality shows the need for expert guidance in risk management, but the complexities in our values also suggest that the expert's analytic techniques for revealing public values may be flawed. Nobody ever said that democracy, with its reliance on the popular will, would produce the best results all of the time, but only results that were preferable on balance to those generated by alternative arrangements. We seem to be more rational than animals, less rational than angels, or computers. In other words, human.

The sources quoted in this article are: Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, vol. 211 (1981); Tversky and Kahneman, "Rational Choice and the Framing of Decisions," *Journal of Business*, vol. 59 (1986); Jack L. Knetsch and J.A. Sinden, "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value," *The Quarterly Journal of Economics* (August 1984); Douglas MacLean, "Risk, Regret, and Rationality," Institute for Philosophy and Public Policy Working Paper RR-5 (forthcoming), and unpublished manuscripts; Paul Slovic, "Preference Reversals," Institute for Philosophy and Public Policy Working Paper RR-2 (March 1987).