PART 3. COGNITIVE CHALLENGES IN ESTIMATING UNCERTAINTY

While our brains are very good at doing many tasks, we do not come hard-wired with statistical processors. Over the past several decades, experimental psychologists have begun to identify and understand a number of the "cognitive heuristics" we use when we make judgments that involve uncertainty.

The first thing to note is that people tend to be systematically overconfident in the face of uncertainty – that is, they produce probability distributions that are much too narrow. Actual values, once they are known, often turn out to lie well outside the tails of their previous distribution. This is well illustrated with the data in the summary table reproduced in Figure 3.1. This table reports results from laboratory studies in which, using a variety of elicitation methods, subjects were asked to produce probability distributions to indicate their estimates of the value of a number of well known quantities. If the respondents were "well calibrated," then the true value of the judged quantities should fall within the 0.25 to 0.75 interval of their probability distribution about half the time. We call the frequency with which the true value actually fell within that interval the interquartile index. Similarly, the frequency with which the true value lies below the 0.01 or above the 0.99 probability values in their distribution is termed the "surprise index." Thus, for a well-calibrated respondent, the surprise index should be 2%.

In these experimental studies, interquartile indices typically were between 20 and 40% rather than the 50% they should have been, and surprise indices ranged from a low of 5% (2.5 times larger than it should have been) to 50% (25 times larger than it should have been).

Overconfidence is not unique to non-technical judgments. Henrion and Fischhoff (1986) have examined the evolution of published estimates of a number of basic physical constants, as compared to the best modern values. Figure 3.2 shows results for the speed of light. While one might expect error bars associated with published experimental results not to include all possible sources of uncertainty, the "recommended values" do attempt to include all uncertainties. Note that for a period of approximately 25 years during the early part of the last century, the one standard deviation error bar being reported for the recommended values did not include the current best estimate.

Three cognitive heuristics are especially relevant in the context of decision making under uncertainty: availability; anchoring and adjustment; and representativeness. For a comprehensive review of much of this literature see Kahneman *et al.* (1982).

When people judge the frequency of an uncertain event they often do so by the ease with which they can recall such events from the past, or imagine such events occurring. This "availability heuristic" serves us well in many situations. For example, if I want to judge the likelihood of encountering a traffic police car on the way to the airport mid-afternoon on a work day, the ease with which I can recall such encounters from the past is probably proportional to the likelihood that I will encounter one today, since I have driven that route many times at that time of day. However, if I wanted to make the same judgment for 3:30 a.m. (a time at which I have never driven to the airport), using availability may not yield a reliable judgment.

A classic illustration of the availability heuristic in action is provided in Figure 3.3A which shows results from a set of experimental studies conducted by Lichtenstein *et al.* (1978) in which well educated Americans were told that 50,000 people die each year in the United States from motor vehicle accidents[21], and were then asked to estimate the number of deaths that occurred each year from a number of other causes. While there is scale compression - the likelihood of high probability events is underestimated by about an order of magnitude, and the likelihood of low probability events is overestimated by a couple orders of magnitude - the fine structure of the results turns out to be replicable, and clearly shows the operation of availability. Many people die of stroke, but the average American hears about such deaths only when a famous person or close relative dies, thus the probability of stroke is underestimated. Botulism poisoning is very rare, but whenever anyone dies the event is covered extensively in the news and we all hear about it. Thus, through the operation of availability, the probability of death from botulism poisoning is overestimated. In short, judgments can be dramatically affected by what gets one's attention. Things that come readily to mind are likely to have a large effect on peoples' probabilistic judgments. Things that do not come readily to mind may be ignored. Or to paraphrase the 14th century proverb, all too often out of sight is out of mind.

We can also illustrate "anchoring and adjustment" with results from a similar experiment in which Lichetenstein *et al.* (1978) made no mention of deaths from motor vehicle accidents but instead told a different group of respondents that about 1000 people die each year in the United States from electrocution. Figure 3.3B shows the resulting trend lines for the two experiments.

[21]Today, while Americans drive more, thanks to safer cars and roads, and reduced tolerance for drunk driving, the number has fallen to about 40,000 deaths per year.

Because in this case respondents started with the much lower "anchor" (1000 rather than 50,000) all their estimates are systematically lower.

One of the most striking experimental demonstrations of anchoring and adjustment was reported by Tversky and Kahneman (1974):

> In a demonstration of the anchoring effect, subjects were asked to estimate various quantities stated in percentages (for example, the percentage of African countries in the United Nations). For each quantity a number between 0 and 100 was determined by spinning a wheel of fortune in the subject's presence. The subjects were instructed to indicate first whether that number was higher or lower than the value of the quantity, and then to estimate the value of the quantity by moving upward or downward from the given quantity. Different groups were given different numbers for each quantity, and these arbitrary numbers had a marked effect on the estimates. For example, the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points[22]. Payoffs for accuracy did not reduce the anchoring effect.

Very similar results are reported for similarly posed questions about other quantities such as "what is the percentage of people in the United States today who are age 55 or older."

The heuristic of "representativeness" says that people expect to see in single instantiations, properties that they know that a process displays in the large. Thus, for example, people judge the sequence of coin tosses HHHTTT to be less likely than the sequence HTHHTH because the former looks less random than the latter, and they know that the process of tossing a fair coin is a random process.

Psychologists refer to feeling and emotion as "affect." Slovic *et al.* (2004) suggest that:

> Perhaps the biases in probability and frequency judgment that have been attributed to the availability heuristic…may be due, at least in part, to affect. Availability may work not

[22]Hastie and Dawes (2001) report that at the time the experiment was conducted the actual value was 35%.

> only through ease of recall or imaginability, but because remembered and imagined images come tagged with affect.

Slovic *et al.* (2004) argue that there are two fundamental ways that people make judgments about risk and uncertainty – one, the "analytic system" the other the "experiential system." They note that while the analytic system "…is rather slow, effortful and requires conscious control," the experiential system is "intuitive, fast, mostly automatic, and not very accessible to conscious awareness." They note that both are subject to various biases and argue both are often needed for good decision making:

> Even such prototypical analytic exercises as proving a mathematical theorem or selecting a move in chess benefit from experiential guidance, the mathematician senses whether the proof "looks good" and the chess master gauges whether a contemplated move "feels right", based upon stored knowledge of a large number of winning patterns. (DeGroot, 1970)

Psychologists working in the general area of risk and decision making under uncertainty are somewhat divided about the role played by emotions and feelings (*i.e.*, affect) in making risk and related judgments. Some (*e.g.*, Sjöberg, 2006) argue that such influences are minor, others (*e.g.*, Loewenstein, 1996; Loewenstein *et al.*, 2001) assign them a dominant role. Agreeing with Slovic *et al.'s* conclusion that both are often important, Wardman (2006) suggests that the most effective responses …"may in fact occur when they are driven by both affective and deliberative-analytical considerations, and that it is the absence of one or the other that may cause problems…"

	Number of assessments N	Interquartile index (ideal 50%)	Surprise index (ideal 2%)
Alpert & Raiffa (1969)			
Group 1-A	880	33	46
Group 2 & 3	1,670	33	39
Group 4	600	36	21
Hession & McCarthy (1974)			
Fractiles	2,035	25	47
Selvidge (1975)			
Five fractiles	400	56	10
Seven fractiles	520	50	7
Schaefer & Borcherding (1973)			
Fractiles	396	23	39
Hypothetical sample	396	16	50
Pickhardt & Wallace (1974)			
Group 1	?	39	32
Group 2	?	30	46
Seaver, von Winterfeldt, & Edwards (1978)			
Fractiles	160	42	34
Odds-fractiles	160	53	24
Probabilities	180	57	5
Odds	180	47	5
Log-odds	140	31	20
Stael von Holstein (1971)			
Fixed intervals	1,269	27	30
Murphy & Winkler (1974 & 1977)			
Fixed intervals	132	45	27 (ideal 25)
Fractiles	432	54	21 (ideal 25)
Schaefer (1976)			
Fixed interval	660	27	25
Lichtenstein & Fischhoff (1978)			
Fractiles	924	33	41
Seaver (1978)			
Parameters of beta dist.	3,200	29	25

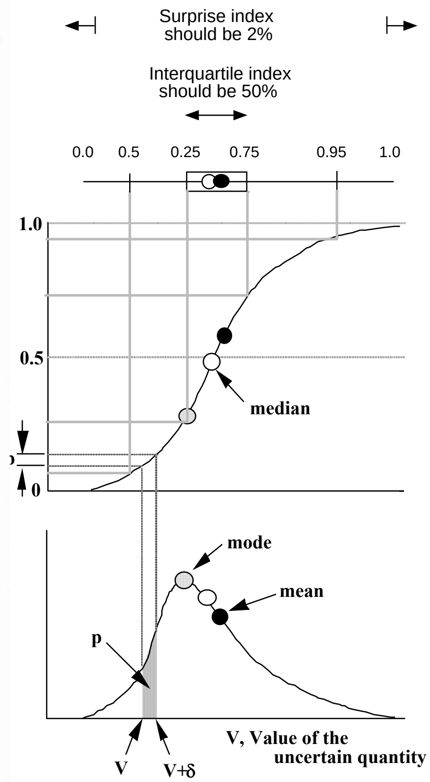

Figure 3.1 Summary of data from different studies in which, using a variety of methods, people were asked to produce probability distributions on the value of well known quantities (such as the distance between two locations), so that their distributions can be subsequently checked against true values. The results clearly demonstrate that people are systematically overconfident (*i.e.*, produce subjective probability distributions that are too narrow) when they make such judgments. The table is reproduced from Morgan and Henrion (1990) who, in compiling it, drew in part on Lichtenstein *et al.* (1982). Definitions of interquartile index and surprise index are shown in the diagram on the right.

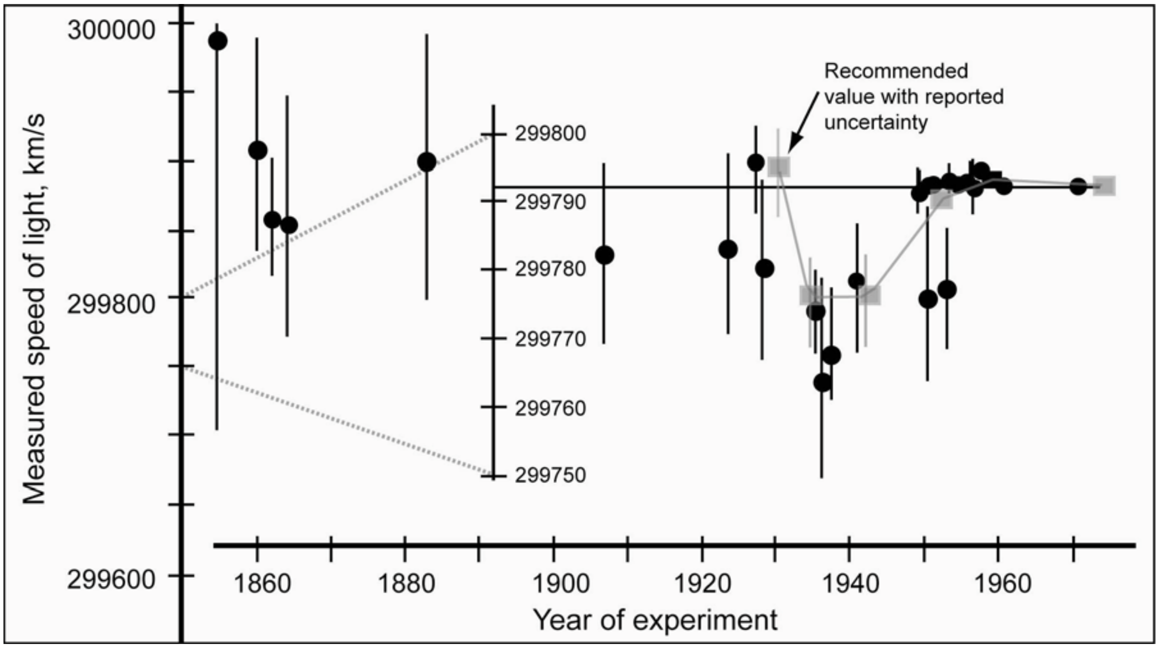

Figure 3.2 Time series of reported experimental values for the speed of light over the period from the mid-1800's to the present (black points). Recommended values are shown in gray. These values should include a subjective consideration of all relevant factors. Note, however, that for a period of approximately 25 years during the early part of the last century, the uncertainty being reported for the recommended values did not include the current best estimate. Similar results obtained for recommended values of other basic physical quantities such as Planck's constant, the charge and mass of the electron and Avogadro's number. For details see Henrion and Fischhoff (1986) from which this figure has been redrawn.

PART 3 REFERENCES

DeGroot, M., 1970: *Optimal Statistical Decision*. McGraw-Hill, New York, 489 pp.

Hastie, R. and R.M. Dawes, 2001: *Rational Choice in an Uncertain World: The psychology of judgment and decision making*. Sage, Thousand Oaks, CA, 372 pp.

Henrion, M. and B. Fischhoff, 1986: Assessing uncertainty in physical constants. *American Journal of Physics*, **54**, 791-798.

Kahneman, D., P. Slovic, and A. Tversky (eds.), 1982: *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, 551 pp.

Lichtenstein, S., P. Slovic, B. Fischhoff, M. Layman, and B. Combs, 1978: Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 551-578.

Lichtenstein, S., B. Fischhoff, and L.D. Phillips, 1982: Calibration of probabilities: The state of the art to 1980. In: *Judgment Under Uncertainty: Heuristics and biases* [Kahneman, D., P. Slovic, and A. Tversky (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, pp. 306-334.

Loewenstein, G.F., 1996: Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, **65**, 272-292.

Loewenstein, G.F., E.U. Weber, C.K. Hsee, and E.S. Welch, 2001: Risk as feelings. *Psychological Bulletin*, **127**, 267-286.

Morgan, M.G. and M. Henrion, 1990: *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, 332 pp.

Sjöberg, L., 2006: Will the real meaning of affect please stand up? *Journal of Risk Research*, ***9***, 101-108.

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**, 311-322.

Tversky, A. and D. Kahneman, 1974: Judgments under uncertainty: Heuristics and biases. *Science*, **185**, 1124-1131.

Wardman, J.K., 2006: Toward a critical discourse on affect and risk perception. *Journal of Risk Research*, **9**, 109-124.

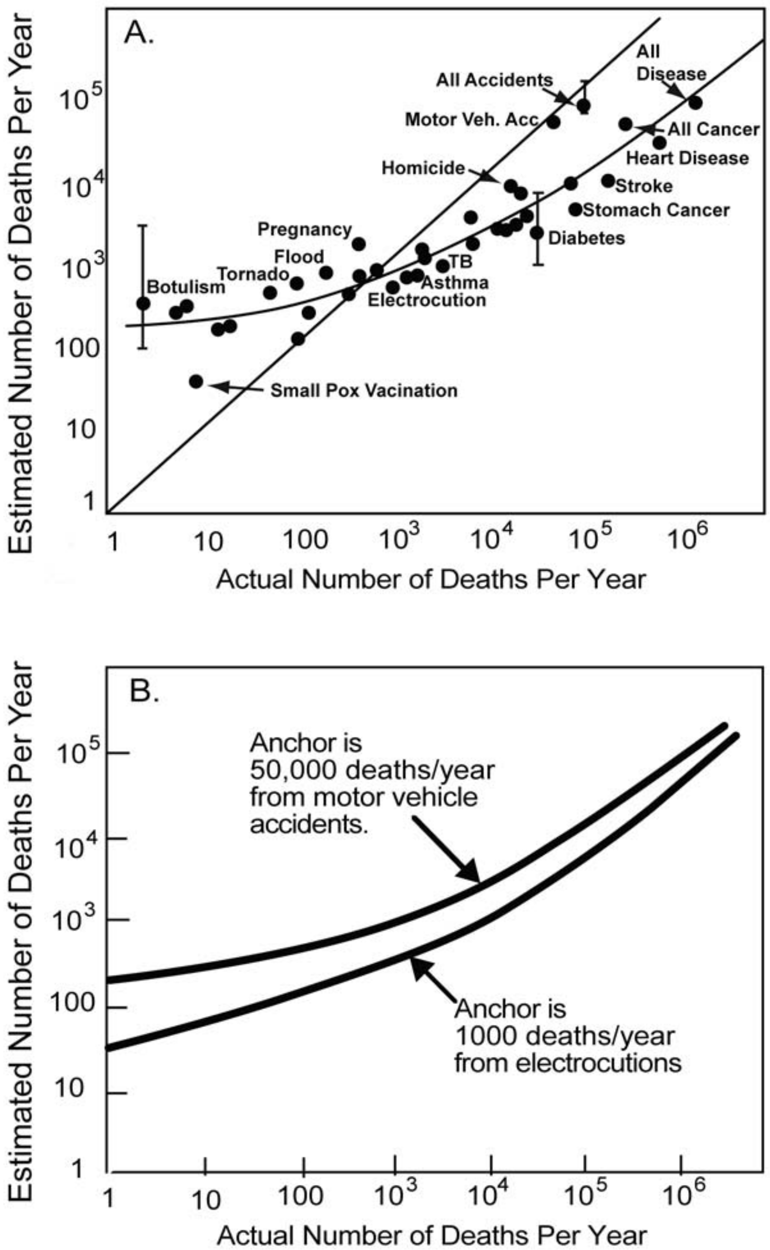

Figure 3.3 Illustration of the heuristic of availability (A) and of anchoring and adjustment (B). In the upper figure, note that stroke lies below the trend line and that botulism lies above the trend line – this is a result of the availability heuristic – we do not learn of most stroke deaths and we do learn of most botulism deaths via news reports. The lower figure replicates the same study with an anchor of 1000 deaths per year. Due to the influence of this lower anchor through the heuristic of anchoring and adjustment, the mean trend line has moved down. Figures are redrawn from Lichtenstein *et al.* (1978).