Journal of Experimental Psychology:
Human Learning and Memory

VOL. 4, NO. 6 NOVEMBER 1978

Judged Frequency of Lethal Events

Sarah Lichtenstein, Paul Slovic, Baruch Fischhoff,
Mark Layman, and Barbara Combs
Decision Research, A Branch of Perceptronics
Eugene, Oregon

A series of experiments studied how people judge the frequency of death from various causes. The judgments exhibited a highly consistent but systematically biased subjective scale of frequency. Two kinds of bias were identified: (a) a tendency to overestimate small frequencies and underestimate larger ones, and (b) a tendency to exaggerate the frequency of some specific causes and to underestimate the frequency of others, at any given level of objective frequency. These biases were traced to a number of possible sources, including disproportionate exposure, memorability, or imaginability of various events. Subjects were unable to correct for these sources of bias when specifically instructed to avoid them. Comparisons with previous laboratory studies are discussed, along with methods for improving frequency judgments and the implications of the present findings for the management of societal hazards.

How well can people estimate the frequencies of the lethal events they may encounter in life (e.g., accidents, diseases, homicides, suicides, etc.)? More specifically, how small a difference in frequency can be reliably detected? Do people have a consistent internal scale of frequency for such events? What factors, besides actual frequency, influence people's judgments?

The answers to these questions may have great importance to society. Citizens must assess risks accurately in order to mobilize society's resources effectively for reducing hazards and treating their victims. Official recognition of the importance of valid risk assessments is found in the "vital statistics" that are carefully tabulated and periodically reported to the public (see Figure 1). There is, however, no guarantee that these statistics are reflected in the public's intuitive judgments.

Few studies have addressed these questions. Most investigations of judged frequency have been laboratory experiments

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contracts N00014-76-C-0074 and N00074-78-C-0100 (ARPA Order Nos. 3052 and 3469) under subcontract to Oregon Research Institute and Subcontracts 76-030-0714 and 78-072-0722 to Perceptronics, Inc. from Decisions and Designs, Inc.

We would like to thank Nancy Collins and Peggy Roecker for extraordinary diligence and patience in typing and data analysis. We are also grateful to Ken Hammond, Jim Shanteau, Amos Tversky, and an anonymous reviewer for perceptive comments on various drafts of this article.

Requests for reprints should be sent to Sarah Lichtenstein, Decision Research, 1201 Oak Street, Eugene, Oregon 97401.

Figure 1. Eat, drink and be merry.

using sequential or simultaneous displays of lights, letters, numbers or horizontal and vertical lines. In such tasks, people's estimates of frequency and proportion have typically been quite accurate. According to Peterson and Beach (1967), the most striking aspect of many of these studies was that the relation between estimated and actual frequency was described well by the identity function. Howell's (1973) review of the literature concluded that "subjects show a remarkable facility for synthesizing and storing the repetitive attribute of event occurrences. They seem capable of maintaining a number of separate frequency streams concurrently as evidenced by the creditable accuracy of frequency retrieval" (p. 51). Similarly, Estes (1976) observed that subjects in probability-learning experiments were "extremely efficient" (p. 51) at acquiring relative-frequency information.

Despite these optimistic conclusions, some studies have found inaccuracies. For example, Attneave (1953) and Hintzman (1969) found that judged frequency increased with the log of the true frequency. Still other studies have suggested some cognitive processes that could lead to even more serious errors in judgments of lethal events. In this regard, Postman (1964) noted that frequency learning is typically incidental learning, which is strongly influenced by selective attention. Estes (1976) observed that accurate learning of frequencies requires the learner to "attend to and encode occurrences of all the alternative events with equal uniformity or efficiency" (p. 53). Underwood (1969) found that items were judged more frequent under conditions of distributed rather than massed practice, and Hintzman (1977) discussed a great deal of evidence showing that apparent frequency of an item increases with greater spacing between its repetitions in a list. Any of these factors could bias judgments about the frequencies of causes of death. Events that capture our attention and "stick in our mind," like homicide, may appear more frequent than they are. Rare events may be overestimated because their appearances are well spread and distinct. Catastrophic (multifatality) events may be overestimated because of their salience or underestimated because of massed presentation.

Tversky and Kahneman (1973) have argued that people judge the probability or frequency of an event by the ease with which relevant instances can be retrieved from memory or imagined. Reliance on memorability and imaginability as a cue for frequency is called the "availability" heuristic. In the context of lethal events, the concept of availability suggests that one's judgments will be influenced not only by direct experience with death and indirect exposure via movies, books, television, newspapers, and the like, but also by memorable characteristics of the different causes of death, such as sensationalism or vividness. Thus we might expect that the frequencies of dramatic events such as cancer, homicide, or multiple-death catastrophes, which tend to be publicized disproportionately, would be overestimated, while the frequencies of "quiet killers" would be underestimated.

In summary, experimental research shows that although people are very good at tracking event frequencies, the potential exists for serious misjudgment. Even without the ambiguity of this conclusion, the implications of these laboratory studies for judgments regarding causes of death would be unclear. Lethal events are emotion-laden stimuli experienced idiosyncratically over the course of a lifetime. Any one person has direct experience with only a few of these events; knowledge about the other events is gained indirectly, from a wide variety of sources. Some of these events occur thousands of times more frequently than others. No laboratory experiments have even approximated these conditions.

Perhaps more relevant are field surveys by several geographers (Burton, Kates, & White, 1978; Kates, 1978; White, 1974; Kates, Note 1). These studies have indicated that (a) people misjudge the hazards posed by floods, earthquakes, hurricanes, and drought; (b) more frequent hazards are estimated more accurately; and (c) accuracy is increased by both the recency of the hazard's last major occurrence and its impact on one's livelihood.

Judgments concerning the probabilities and frequencies of real-life events have also been studied by Selvidge (1972). In one phase of her research, five subjects first ranked several sets of accidents and crimes according to frequency and then estimated the absolute frequencies. Although her subjects were fairly good at ordering the events, they did a poor job of assigning absolute frequencies. She also found a great amount of variability across subjects, event categories, and response modes. This variability and her small sample size led Selvidge to advocate that these issues be investigated on a much larger scale. The present study does this.

Five experiments are reported here. The first two examine the accuracy of comparative judgments, using a paired-comparison format. The third evaluates judgments of absolute frequency. The fourth examines the role that several aspects of availability may play in determining such judgments. The fifth explores the degree to which subjects can overcome their errors when informed of the nature of their biases.

Table 1
Master List of Causes of Death

Cause	Rate/10^8
Smallpox	0
Poisoning by vitamins	.5
Botulism	1
Measles	2.4
Fireworks	3
Smallpox vaccination	4
Whooping cough	7.2
Polio	8.3
Venomous bite or sting	23.5
Tornado	44
Lightning	52
Nonvenomous animal	63
Flood	100
Excess cold	163
Syphilis	200
Pregnancy, childbirth, and abortion	220
Infectious hepatitis	330
Appendicitis	440
Electrocution	500
Motor vehicle–train collision	740
Asthma	920
Firearm accident	1,100
Poisoning by solid or liquid	1,250
Tuberculosis	1,800
Fire and flames	3,600
Drowning	3,600
Leukemia	7,100
Accidental falls	8,500
Homicide	9,200
Emphysema	10,600
Suicide	12,000
Breast cancer	15,200
Diabetes	19,000
Motor vehicles (car, truck, or bus) accidents	27,000
Lung cancer	37,000
Cancer of the digestive system	46,600
All accidents	55,000
Stroke	102,000
All cancer	160,000
Heart disease	360,000
All disease	849,000

Experiment 1: Paired-Comparison Judgments of Lethal Events

The first experiment investigated the accuracy of relative-frequency judgments for various causes of death.

Method

Stimuli. Table 1 shows the stimulus events, 41 causes of death, and gives, for each item, the frequency of death per 10^8 United States residents per year, based on reports prepared by the National Center for Health Statistics for the years 1968–1973.[1] These events were chosen to represent the range of frequencies of causes of death for which yearly statistics are available. Obscure or unfamiliar causes were excluded, as were causes showing large fluctuations from year to year. For the few chosen events that showed a systematic trend across years (e.g., homicide, which increased from 7,300 per 10^8 in 1968 to 9,400 per 10^8 in 1973), the average over the last 2 years was used.

From these 41 causes of death, 106 pairs were constructed such that (a) each cause appeared in approximately six pairs and (b) the ratios of relative frequencies (comparing the more to the less frequent cause of death) varied systematically from 1.25:1 (example: fireworks vs. measles) to about 100,000:1 (example: stroke vs. botulism). Five pairs included smallpox as the less frequent cause of death. Since no one in the United States has died of smallpox since 1949, the rate shown in Table 1 is zero, and no ratio comparing any other disease with smallpox can be defined. In the results that follow, all analyses employing ratios of true frequencies (called *true ratios*) exclude the five pairs involving smallpox.

Subjects. Two groups of subjects participated. The first, hereafter referred to as the *college students*, consisted of 51 males and 60 females who answered an ad in the University of Oregon campus newspaper. The second consisted of 77 female members of the Eugene, Oregon chapter of the League of Women Voters, a group representative of the best-informed citizens in the community. All subjects were paid for participating. The data were collected from the students in the autumn of 1974 and from the league members in the spring of 1975.

The order of the 106 pairs and of the two causes within each pair was determined randomly. All subjects saw the same random order.

Instructions. The subjects' instructions read as follows:

> Each item in part one consists of two different possible causes of death. The question you are to answer is: Which cause of death is more likely? We do not mean more likely *for you*, we mean more likely *in general*, in the United States.
>
> Consider all the people now living in the United States—children, adults, *everyone*. Now supposing we randomly picked just one of those people. Will that person more likely die next year from cause A or cause B? For example: Dying in a bicycle accident versus dying from an overdose of heroin. Death from each cause is remotely possible. Our question is, which of these two is the more likely cause of death?
>
> For each pair of possible causes of death, A and B, we want you to mark on your answer sheet which cause you think is MORE LIKELY.
>
> Next, we want you to decide how many times more likely this cause of death is, as compared with the other cause of death given in the same item. The pairs we use vary widely in their relative likelihood. For one pair, you may think that the two causes are equally likely. If so, you should write the number 1 in the space provided for that pair. Or, you may think that one cause of death is 10 times, or 100 times, or even a million times as likely as the other cause of death. You have to decide: How many *times* as likely is the more likely cause of death? Write the number in the space provided. If you think it's twice as likely, write 2. If it's 10 thousand times as likely, write 10,000, and so forth.

In the instructions and at the top of the answer sheet we drew a logarithmic scale labeled with both numbers and words for all powers of 10 from 1 to 1,000,000. The scale ended in an arrow to indicate that the scale continued. The instructions continued:

> The scale is there to give you an idea of the kinds of numbers you might want to use. You don't have to use exactly those numbers. You could write 75 if you think that the more likely cause of death is 75 times more likely than the other cause, or 500, if you think that the more likely cause of death is 500 times more likely than the other.
>
> For some pairs, you may believe that one cause of death is just a little bit more likely than the other cause of death. For this situation, you will have to use a decimal point in your answer:
>
> 1.1 means that the more likely cause is 10% more likely than the other cause.
> 1.2 means 20% more likely.
> 1.5 means 50% more likely, or half again as likely.
> 1.8 means 80% more likely.
> 2 means twice as likely, which is the same as 100% more likely.
> 2.5 means two and a half times as likely.

In addition, the following glossary was provided to insure that the subjects understood what was included in some possibly ambiguous categories:

> All accidents: *includes* any kind of accidental event; *excludes* diseases and natural disasters (floods, tornadoes, etc.).
> All cancer: includes leukemia.
> Cancer of the digestive system: includes cancer of stomach, alimentary tract, esophagus, and intestines.

[1] For convenience, these frequencies are referred to in this article as the *true frequencies*, although we recognize that they are statistical estimates.

Excess cold: freezing to death or death by exposure.
Nonvenomous animal: dogs, bears, etc.
Venomous bite or sting: caused by snakes, bees, wasps, etc.

Results

Accuracy. Two measures were computed for each pair of causes of death: the percentage of subjects who correctly selected the more likely item and the geometric mean of the subjects' ratio judgments. For any subject who did not correctly select the more likely cause of death, the inverse of the judged ratio was used in calculating the geometric mean. For example, death by fireworks is more frequent than death from measles. If a subject said measles was 5 times more likely to cause death than fireworks, the inverse, .2, was used. The two summary measures, percentage correct and the geometric mean of the ratio judgments, are shown for all 106 pairs for both groups of subjects in Table 2.

Examination of Table 2 illustrates the many, often severe, misconceptions held by both the college students and the league members. For example, even though stroke causes 85% more deaths than all accidents combined (pair 37, true ratio = 1.85), only 20% of the students and 23% of the league members judged stroke to be more likely. The geometric mean of the ratio judgments was only .04 for the students, indicating that on the average, they believed that accidents were 25 times (1 ÷ .04) more frequent. Tornadoes were seen by the student subjects as more frequent killers than asthma, even though the latter is 20 times more likely (pair 61). Death by lightning was judged less likely than death by botulism even though it is 52 times more frequent (pair 71). Death by asthma was judged only slightly more frequent than death by botulism (pair 91), even though it is over 900 times more frequent! Accidental deaths were reported by the students to be about as likely as death from disease despite a true ratio of 15.4 for diseases over accidents (pair 69).

Some errors were in the opposite direction: A large percentage of subjects knew which cause of death was more likely, but the ratios given were far too large. For example, death by a motor vehicle accident is only 1.4 times more likely than death from diabetes (pair 25), not 356 times more likely (the students' geometric mean) or 100 times more likely (league members).

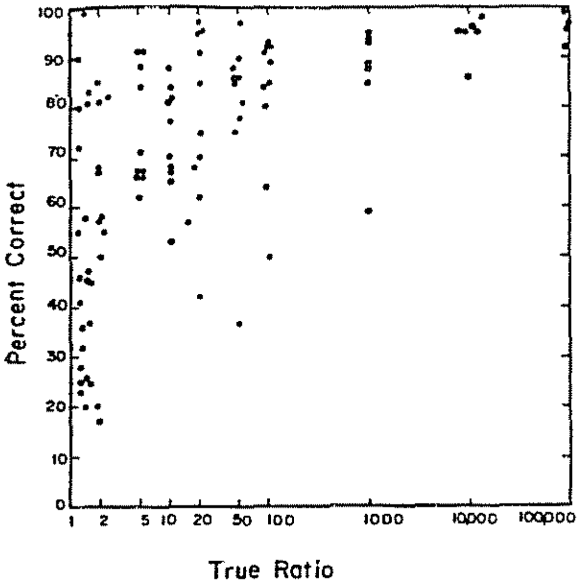

Figure 2. Percentage of student subjects who correctly identified the more likely cause of death as a function of true ratio for 101 paired causes of death.

Subjects' ability to detect the more likely event was not quite as bad as these examples suggest. They were generally able to identify the more frequent cause of death when the true ratio was 2:1 or greater. Below 2:1, however, discrimination was often poor, as shown in Figures 2 and 3, which compare the percentage of correct discriminations with the log true ratio for the two groups of subjects (101 pairs, excluding smallpox).

Accuracy as measured by percentage correct was slightly higher for events with greater true frequency. The partial correlation between percentage correct and log frequency of the less likely event, holding true ratio constant, was .24 ($z = 2.48$, one-tailed $p < .01$) for the college students and .19 ($z = 1.62$, one-tailed $p < .06$) for the league members. Since greater true frequency typically implies greater exposure, these are surprisingly low correlations. They

Table 2
Results of Paired-Comparison Judgments for Causes of Death

Pair no.	Less likely	More likely	Rate of less likely	True ratio	% correct		Geometric mean	
					Students	LWV	Students	LWV
1	Smallpox	Botulism	0	—	77	88	5.54	18.3
2	Smallpox	Measles	0	—	60	86	1.52	16.0
3	Smallpox	Smallpox vaccination	0	—	26	56	.08	1.08
4	Smallpox	Polio	0	—	64	74	2.78	5.90
5	Smallpox	Heart disease	0	—	97	97	1,130.0	13,263
6	Measles	Fireworks	2.4	1.25	28	35	.27	.44
7	Fireworks	Smallpox vaccination	3	1.33	36	29	.47	.29
8	Lightning	Nonvenomous animal	52	1.21	55	58	.71	1.89
9	Excess cold	Syphilis	163	1.23	72	81	5.93	23.7
10	Asthma	Firearm accident	920	1.20	80	78	11.0	14.1
11	Leukemia	Accidental falls	7,100	1.20	46	64	.78	2.45
12	Accidental falls	Emphysema	8,500	1.25	41	39	.65	.45
13	Homicide	Suicide	9,200	1.30	32	30	.19	.14
14	Breast cancer	Diabetes	15,200	1.25	23	27	.13	.14
15	Lung cancer	Stomach cancer	37,000	1.25	25	44	.31	.81
16	Stomach cancer	All accidents	46,600	1.19	90	87	28.1	29.4
17	Tornado	Nonvenomous animal	44	1.43	20	21	.10	.15
18	Flood	Excess cold	100	1.63	37	32	.51	.34
19	Hepatitis	Electrocution	330	1.52	45	38	.61	.44
20	Electrocution	Motor-train collision	500	1.48	45	52	1.16	2.40
21	Poisoning	Tuberculosis	1,250	1.44	26	12	.19	.03
22	Leukemia	Emphysema	7,100	1.49	47	47	.58	1.02
23	Accidental falls	Suicide	8,500	1.41	58	51	1.58	.64
24	Suicide	Diabetes	12,000	1.58	25	31	.09	.20
25	Diabetes	Motor vehicle accident	19,000	1.42	99	96	356.0	99.6
26	Lung cancer	All accidents	37,000	1.49	81	90	11.1	34.7
27	Stroke	All cancer	102,000	1.57	83	75	21.0	8.20
28	Poisoning by vitamins	Botulism	.5	2.00	68	82	4.08	14.0
29	Lightning	Flood	52	1.92	85	87	18.6	14.1
30	Flood	Syphilis	100	2.00	57	73	1.74	6.60
31	Pregnancy, etc.	Appendicitis	220	2.00	17	10	.10	.07
32	Appendicitis	Asthma	440	2.09	50	71	1.00	6.65
33	Tuberculosis	Fire and flames	1,800	2.00	81	92	10.5	38.9
34	Tuberculosis	Drowning	1,800	2.00	67	83	2.98	19.3
35	Leukemia	Breast cancer	7,100	2.14	58	60	1.48	2.98
36	Breast cancer	Lung cancer	15,200	2.43	82	71	6.42	2.74
37	All accidents	Stroke	55,000	1.85	20	23	.04	.13
38	All cancer	Heart disease	160,000	2.25	55	68	.89	1.88
39	Poisoning by vitamins	Measles	.5	4.80	66	74	2.50	5.45
40	Polio	Tornado	8.3	5.30	71	86	4.26	18.0
41	Nonvenomous animal	Hepatitis	63	5.24	84	90	9.76	15.1
42	Syphilis	Firearm accident	200	5.50	66	77	4.38	16.6
43	Pregnancy, etc.	Firearm accident	220	5.00	62	68	2.14	6.66
44	Motor-train collision	Fire and flames	740	4.86	67	90	2.45	14.2
45	Motor-train collision	Drowning	740	4.86	67	77	1.97	7.17
46	Tuberculosis	Homicide	1,800	5.11	91	91	25.2	72.9
47	Emphysema	All accidents	10,600	5.19	88	95	269	107
48	Diabetes	Stroke	19,000	5.37	91	84	29.3	46.6
49	Measles	Venomous bite/sting	2.4	9.79	65	68	1.68	3.11
50	Smallpox vaccination	Tornado	4	11.00	84	91	16.8	22.7
51	Venomous bite/sting	Pregnancy, etc.	23.5	9.36	77	82	7.27	11.6
52	Lightning	Electrocution	52	9.62	88	92	23.1	26.6
53	Hepatitis	Fire and flames	330	10.90	67	78	4.22	11.2
54	Hepatitis	Drowning	330	10.90	68	71	4.55	7.80
55	Poisoning	Suicide	1,250	9.60	70	73	5.50	2.59
56	Tuberculosis	Diabetes	1,800	10.56	53	90	1.47	14.7
57	Emphysema	Stroke	10,600	9.62	81	74	10.5	10.2
58	Lung cancer	Heart disease	37,000	9.73	81	95	3.71	24.4
59	Measles	Tornado	2.4	18.3	68	70	5.63	4.67
60	Polio	Excess cold	8.3	19.6	63	79	1.64	6.05
61	Tornado	Asthma	44	20.9	42	68	.36	3.53
62	Nonvenomous animal	Poisoning	63	19.8	95	95	17.1	64.9
63	Hepatitis	Leukemia	330	21.5	75	79	12.5	14.7
64	Appendicitis	Homicide	440	20.9	91	97	72.7	105
65	Motor-train collision	Breast cancer	740	20.5	70	83	4.90	20.9
66	Poisoning	Motor vehicle accident	1,250	21.6	95	94	388	304
67	Tuberculosis	Lung cancer	1,800	20.6	85	99	23.3	145
68	Diabetes	Heart diasese	19,000	18.9	97	97	127	206
69	All accidents	All diseases	55,000	15.4	57	79	1.62	11.6
70	Poisoning by vitamins	Venomous bite/sting	.5	47.0	75	79	4.04	16.5
71	Botulism	Lightning	1	52.0	37	45	.30	.32
72	Whooping cough	Hepatitis	7.2	45.8	88	91	7.66	12.4
73	Venomous bite/sting	Firearm accident	23.5	46.8	85	87	9.94	35.7

Table 2 (*continued*)

Pair no.	Less likely	More likely	Rate of less likely	True ratio	% correct: Students	% correct: LWV	Geometric mean: Students	Geometric mean: LWV
74	Venomous bite/sting	Poisoning	23.5	53.2	78	83	5.21	11.3
75	Syphilis	Homicide	200	46.0	86	82	31.7	44.2
76	Pregnancy, etc.	Suicide	220	54.5	81	79	14.8	12.7
77	Electrocution	Motor vehicle accident	500	54.0	97	99	539	909
78	Asthma	Stomach cancer	920	50.4	86	90	36.4	43.4
79	Leukemia	Heart disease	7,100	50.7	90	99	33.8	184
80	Poisoning by vitamins	Lightning	.5	104	50	61	1.45	2.00
81	Measles	Pregnancy, etc.	2.4	92	84	79	13.4	9.44
82	Whooping cough	Motor-train collision	7.2	103	85	95	15.2	70.1
83	Flood	Homicide	100	92	91	94	81.7	294
84	Excess cold	Breast cancer	163	93	80	95	26.7	255
85	Syphilis	Diabetes	200	94	64	71	2.36	4.52
86	Appendicitis	Stomach cancer	440	105	89	95	30.2	83.7
87	Drowning	Heart disease	3,600	100	92	95	56.9	272
88	Fire and flames	Heart disease	3,600	100	93	94	79.7	252
89	Accidental falls	All disease	8,500	100	91	90	324	614
90	Poisoning by vitamins	Electrocution	.5	1,000	85	86	30.7	27.7
91	Botulism	Asthma	1	920	59	75	1.50	9.29
92	Botulism	Firearm accident	1	1,100	88	94	29.9	189
93	Whooping cough	Leukemia	7.2	983	94	99	45.2	166
94	Polio	Accidental falls	8.3	1,024	93	95	17.5	198
95	Flood	Stroke	100	1,020	89	90	43.4	186
96	Excess cold	All cancer	163	982	95	99	1,490	4,337
97	Botulism	Emphysema	1	10,600	86	94	24	190
98	Measles	Motor vehicle accident	2.4	11,250	96	99	1,070	2,765
99	Fireworks	Motor vehicle accident	3	9,000	95	97	1,430	5,268
100	Polio	Stroke	8.3	12,289	95	99	164	2,265
101	Tornado	Heart disease	44	8,182	95	97	348	2,396
102	Nonvenomous animal	All disease	63	13,476	98	100	5,600	33,521
103	Poisoning by vitamins	Stomach cancer	.5	92,800	92	99	103	578
104	Botulism	Stroke	1	102,000	96	95	106	468
105	Fireworks	Heart disease	3	120,000	96	99	1,530	5,779
106	Smallpox vaccination	Heart disease	4	90,000	99	100	3,610	12,244

Note. LWV = subjects who were members of the League of Women Voters.

reflect the fact that our subjects were not much better at judging the relative frequency of death from high-frequency pairs such as all cancer versus heart disease (pair 38) than they were at judging low-frequency pairs such as poisoning by vitamins versus botulism (pair 28).

The geometric means of the likelihood judgments were only moderately related to the true ratios of frequencies, as shown in Figures 4 and 5 (101 pairs, excluding smallpox). For example, the college students produced mean ratios in the range of 100:1 to 500:1 for pairs with true ratios as small as 1.5:1 and as large as 100,000:1! Conversely, pairs having true ratios of about 2:1 had geometric mean judgments ranging from 25:1 in the wrong direction to over 300:1 in the right direction! The geometric means were somewhat more accurate for the league members but still were far from optimal. The correlation between log geometric mean judged ratio and log true ratio was .69 for the students and .75 for the league members. The regression lines (shown as dashed lines in Figures 4 and 5) were both too flat.

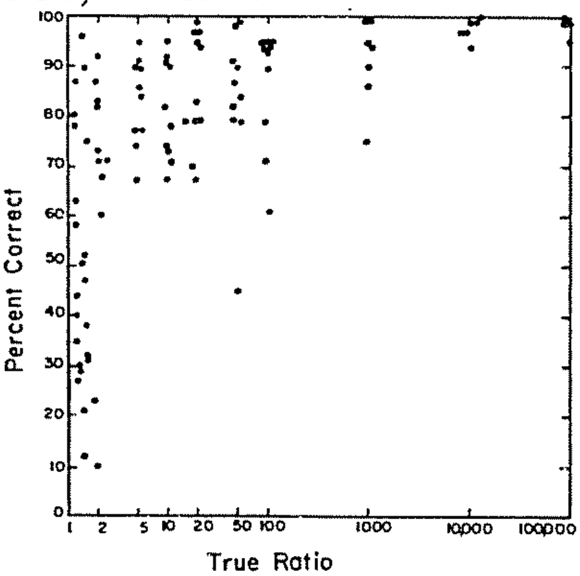

Figure 3. Percentage of League of Women Voters who correctly identified the more likely cause of death as a function of true ratio for 101 paired causes of death.

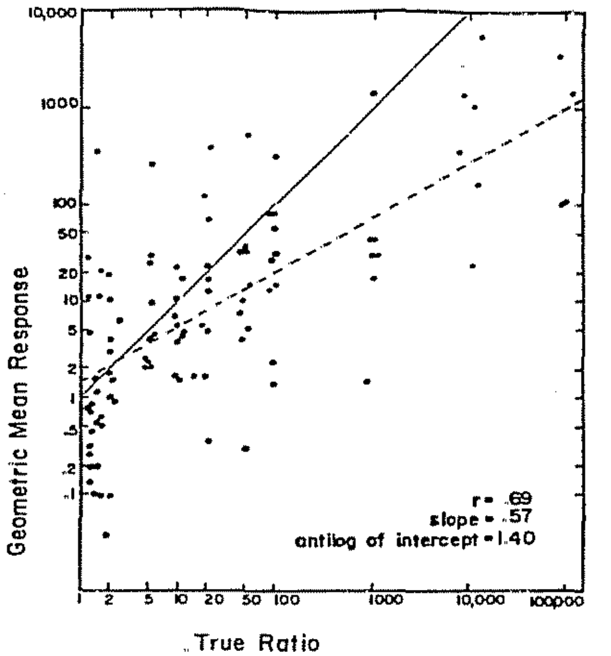

Figure 4. Geometric means of student subjects' ratio judgments as a function of true ratio for 101 paired causes of death.

Secondary bias. The regression lines shown in Figures 4 and 5 capture what we will call *primary bias:* a tendency to underestimate large ratios. In addition, the data showed a *secondary bias:* Different pairs with the same ratio had quite different judged ratios. One measure of this secondary bias is the signed difference between the log geometric mean for a pair and its log geometric mean as predicted by the regression equation. (This measure is equivalent to the vertical distance between a point in Figure 4 or 5 and the dashed regression line.) A positive value indicates that the ratio judgments for that pair were large relative to the general relationship between the judged ratio and the true ratio. A negative value indicates relative underestimation or estimation in the wrong direction. As measured by these residual values, secondary bias was highly consistent across the two groups of subjects: The between-groups correlation of the residuals was .90 (over 101 pairs). Further analysis of secondary bias will be presented later in the article.

Consistency. Even though they were often inaccurate, subjects' mean responses revealed a consistent subjective ordering for the causes of death. There were 18 triads (involving 29 of the 41 causes of death) of the form *A vs. B, B vs. C, A vs. C* within the 106 pairs (for example, *all accidents* paired with *stroke, stroke* paired with *emphysema,* and *emphysema* paired with *all accidents*). For such triads, we asked *Were the choice percentages transitive?* and *Were the geometric means consistent?* The answer to both these questions was yes for the triad described above. Eighty percent of the students said *all accidents* were more likely than *stroke* (geometric mean likelihood ratio = 26.3), 81% said *stroke* was more likely than *emphysema* (geom. mean likelihood ratio = 10.5), and 88% said *all accidents* were more likely than *emphysema* (geom. mean likelihood ratio = 269). These data exhibit strong stochastic transitivity,[2] in that the percentage of subjects judging *all accidents* more likely than *emphysema* was the largest of the three percentages. The consistency of the geometric means is shown by the similarity of the third mean (269) to the product of the first two means (276). Thus, the group showed a clear subjective ordering: *emphysema* < *stroke* < *all accidents*. The true order, however, is *emphysema* < *all accidents* < *stroke*. These results are typical of all 36 triads analyzed (18 triads each for college students and league members). The choice percentages exhibited weak stochastic transitivity for every triad; strong stochastic transitivity was satisfied for 27 out of 36 triads.

The consistency of the ratio judgments was measured by comparing the log of the geometric mean ratio for pair A:C in each

[2] Three levels of stochastic transitivity may be distinguished (cf. Coombs, Dawes, & Tversky, 1970, p. 156). For any three stimuli, x, y, and z, assume that $p(x, y) \geq 1/2$ (i.e., that the proportion choosing x over y is greater than or equal to .5) and that $p(y, z) \geq 1/2$. Then *strong* stochastic transitivity requires that $p(x, z) \geq \max[p(x, y), p(y, z)]$, *moderate* stochastic transitivity requires that $p(x, z) \geq \min[p(x, y), p(y, z)]$, while weak stochastic transitivity requires only that $p(x, z) \geq 1/2$.

triad with the log of the product of the geometric mean ratios for A:B and B:C. The relationship was linear with $r = .99$ (slope $= 1.10$; antilog of intercept $= .83$) for the college students and $r = .97$ (slope $= 1.05$; antilog of intercept $= 1.09$) for the league members. These results suggest that as a group, these subjects exhibited an interval scale of subjective frequency.

Between-groups comparisons. The responses of the students and the league members were highly similar. Across all 106 pairs, the correlation between the two groups was .93 for both percentage correct and geometric mean judged ratio. The high correlation between the two groups' secondary bias residuals is further evidence of this similarity. The league members had a somewhat higher percentage correct than the students ($M = 76.8$ vs. 71.3); their percentage correct was higher for 80 pairs, equal for 5 pairs, and lower for 21 pairs (sign test; $p < .001$). For the ratio judgments, however, the league members did not perform significantly better than the students; the geometric mean of their ratio judgments was closer to the true ratio for only 62 of the 106 pairs (sign test; $z = 1.65$, $p > .10$).

Individual performance. The performance of individual subjects was rather variable. Percent correct ranged between 56% and 84% for the students and between 60% and 89% for the league members. Analysis of the correlations between log judged ratio and log true ratio over 101 items indicated that few individuals showed any appreciable ability to perform the ratio-estimation task. These correlations ranged between −.11 and .72 ($Mdn = .45$) for students and between .10 and .80 ($Mdn = .51$) for league members.

Further insight into the level of individual subjects' performance was obtained by calculating an error ratio, defined as the ratio of the judgment to the truth, or vice versa, whichever was greater than 1. A subject who always gave a judged ratio off by a factor of 10, that is, either 10 times as large or a tenth as large as the true ratio, would have a mean error ratio of 10. The median student subject erred by a factor of 22.5

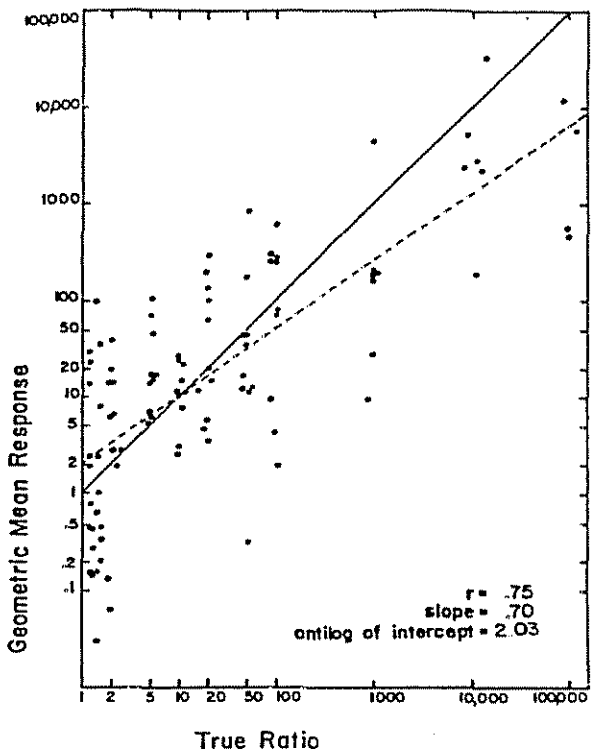

Figure 5. Geometric means of League of Women Voters subjects' ratio judgments as a function of true ratio for 101 paired causes of death.

(range = 7–556), while the median league member erred, on the average, by a factor of 17.6 (range = 5–2,693).

An analysis of transitive triads was also done separately for each subject. The median number of transitive triads (out of 18) was 17 for each group. Only 27% of the subjects in both groups showed more than one intransitivity, while 44% (students) and 49% (league members) were always transitive. Thus, the strong internal consistency found in the group means was also found in the judgments of individuals.

Experiment 2: Paired-Comparison Judgments of Words and Occupations

In order to test whether the primary results of Experiment 1 were unique to the set of stimuli used, Experiment 1 was repeated using pairs of words and pairs of occupations as stimuli.

Method

Stimuli. The list of words studied is shown in Table 3, along with their frequency of occurrence per 10^6 words of English text. These frequencies represent an average from two separate sources. One source, the Lorge magazine count (Thorndike & Lorge, 1944), analyzed frequencies from a sample of about a million words from each of five major magazines between the years 1927 and 1938. The second source (Kučera & Francis, 1967) analyzed 500 samples of about 200 words each, taken from a wide variety of materials, ranging from newspapers to scientific journals and from popular romantic fiction to abstruse philosophical discussions. For the words in Table 3, the frequencies estimated by the two sources agreed closely. From this list, 100 pairs of words were selected, with true ratios ranging from 1.10 (*of* vs. *to*) to 6,126 (*the* vs. *cork*).

The list of occupations studied is shown in Table 4, along with their frequency of occurrence among 10^6 employed U.S. civilian citizens. These frequencies were derived from a report compiled by the U.S. Bureau of the Census (1972). From the list, 95 pairs were selected, with true ratios ranging from 1.15 (garbage collector vs. upholsterer) to 1,229 (registered nurse vs. lay midwife).

Subjects and instructions. The subjects were college students recruited via a campus newspaper advertisement and paid for their participation. A group of 111 subjects judged the word pairs, and a different group of 118 individuals judged occupations. The instructions for words and occupations paralleled those for causes of death. For pairs of words, the subjects were asked to judge which word is more likely to be sampled at random from common writing (magazines and books, fiction, nonfiction, scientific, nonscientific, etc.) in the United States, and to indicate how many times more likely the more frequent word is than the other word in the pair. For occupations, subjects were asked to indicate whether an employed U.S. citizen picked at random is more likely to be working as an A or a B, and how many times more likely the more frequent occupation is than the other occupation in the pair.

Table 3
Master List of Words

Word	Rate/10^6
the	61,260
of	34,716
and	29,834
to	25,892
in	19,032
that	11,483
he	10,246
for	9,118
with	7,300
on	6,730
from	4,044
when	2,807
out	2,565
time	1,751
two	1,368
after	1,152
people	821
again	730
once	578
next	455
half	358
result	222
music	182
couple	125
hit	104
proud	70
dull	46
tent	26
cork	10
jug	7
bun	2

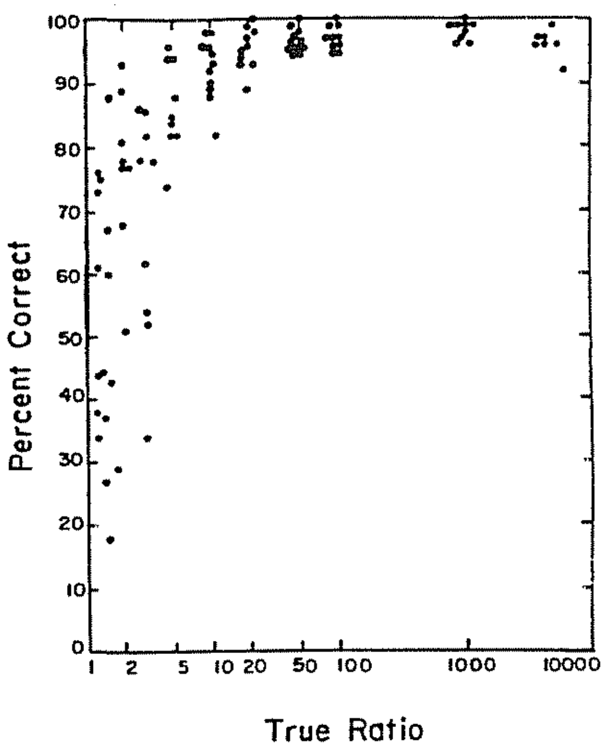

Figure 6. Percentage of subjects who correctly identified the more likely cause of death as a function of true ratio for 100 pairs of words.

Results

Accuracy. Figures 6 and 7 show the relationship between percentage correct and true ratio, and geometric mean ratio judgments are plotted against true ratio in Figures 8 and 9.[3]

For true ratios of 5:1 or greater, percentage correct was considerably higher for

[3] The tables on which these figures are based may be obtained from the authors.

words than for occupations; below 5:1 there was no difference. For true ratios larger than 2:1, both words and occupations were more accurately discriminated than were causes of death (compare Figures 6 and 7 with Figure 2). For true ratios $\leq 2{:}1$, there were again numerous errors of discrimination.

Geometric mean judged ratios for words and occupations were considerably closer to the corresponding true ratios than were judged ratios for causes of death, as may be seen by comparing Figures 8 and 9 (words and occupations) with Figure 4 (causes of death). The correlation between judged and true ratios was higher for words (.90) than for occupations (.81), but since the scatter about the regression line is not notably less, this effect may be attributed to the greater range of true ratios for words. As shown in Figures 8 and 9, the slope of the regression line for occupations was somewhat flat, but words showed a slope near unity, which, taken with the intercept of 1.95, indicated a systematic tendency toward overestimation.[4]

Consistency. The consistency of subjective ordering of the stimuli was sought by analyzing the triads in the words and occupations pairs. Of the 39 triads contained in the words task, 28 showed strong stochastic transitivity, 10 showed moderate stochastic transitivity, and one was intransitive. The one intransitive triad involved three pairs for which consensus was lacking

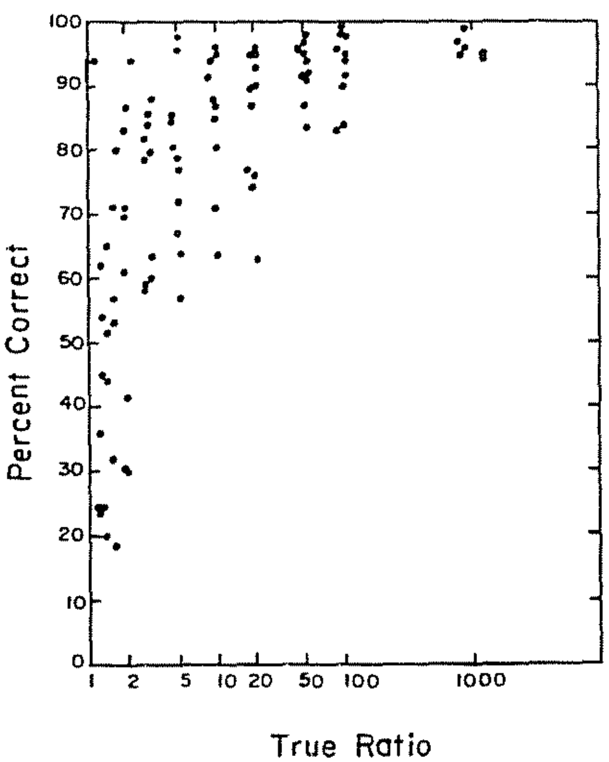

Figure 7. Percentage of subjects who correctly identified the more likely cause of death as a function of true ratio for 95 pairs of occupations.

Table 4
Master List of Occupations

Occupation	Rate/10[8]
Secretary	3,529,680
Elementary or secondary school teacher	3,155,206
Retail sales clerk	2,967,880
Truck driver	1,802,169
Waiter or waitress	1,331,616
Registered nurse	1,083,800
Auto mechanic	1,051,250
College or university teacher	635,138
Electrician	611,935
Telephone operator	531,655
Physician	436,322
Lawyer	339,829
Letter carrier	329,866
Bus driver	308,205
Bartender	246,584
Computer programmer	210,750
Librarian	159,172
Baker	142,634
Bulldozer operator	115,537
Garbage collector	93,290
Upholsterer	81,118
Architect	73,418
Dietitian	52,422
Airline purser, steward, or stewardess	43,891
Air traffic controller	33,040
Airline pilot or copilot	32,787
Psychiatrist	28,191
Veterinarian	25,387
Motion picture projectionist	20,198
Judge	16,001
FBI special agent	10,320
Rabbi	8,491
Embalmer	6,203
EEG technician	3,919
Jockey	2,065
Nuclear reactor operator	1,568
Lay midwife	882

[4] Carroll (1971), who elicited direct (magnitude) estimates of 60 words (12 of which were used here), found a correlation of .92 between assessed and actual values. His regression line had a slope of .58.

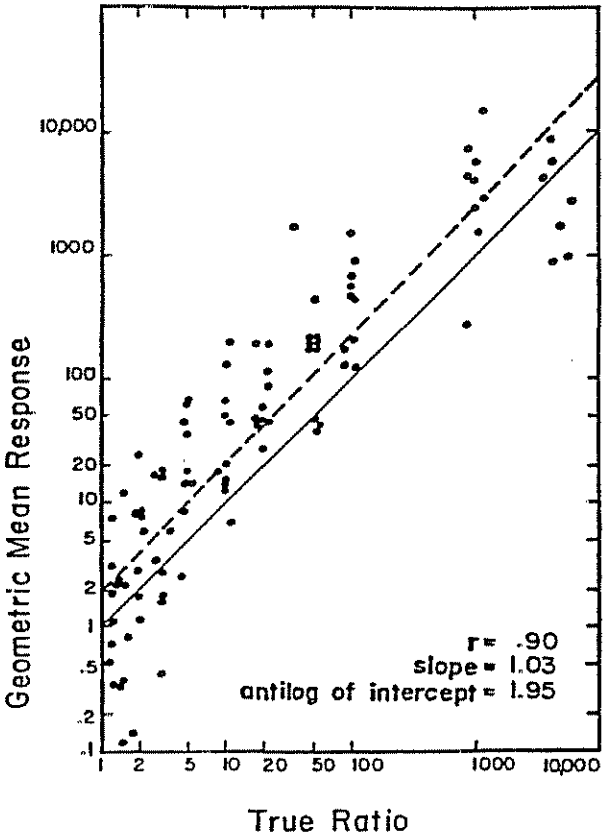

Figure 8. Geometric means of subjects' ratio judgments as a function of true ratio for 100 pairs of words.

(57% of subjects thought *in* was more likely than *that*; 56%, *that* more likely than *for*; and 51%, *for* more likely than *in*). Of the 20 triads contained in the occupations task, 17 showed strong stochastic transitivity, and 3 showed moderate stochastic transitivity.

The log geometric mean ratio response to the third pair of each triad was correlated with the log of the product of the responses of the other two pairs; these correlations were .94 for words (slope = 1.21, antilog of intercept = .80) and .76 for occupations (slope = .64, antilog of intercept = 5.32). Thus, words and occupations judgments showed considerable internal consistency, as found with causes of death.

Comparison with Experiment 1. The purpose of Experiment 2 was to find out whether the major findings of Experiment 1 were specific to lethal events. Three results of this comparison are noteworthy. First, subjects responded more accurately to words than to occupations; causes of death were worse yet. This may be due to exposure: We experience many more samples of English text each day than examples of people working in occupations, and our exposure to death is even more limited. Another possible reason for poorer performance with causes of death is that our exposure to these events is systematically biased. We shall discuss this bias later in the article.

Second, we found that causes-of-death subjects tended to underestimate ratios larger than 50:1. Underestimation did not appear at all with words and was found with occupations only for ratios of 1,000:1. Thus, one cannot conclude that the primary bias found in Experiment 1 was simply due to difficulties in using large numbers rather than to insufficient discrimination between different causes of death.

Third, we found strong evidence in these new tasks that subjects possess consistent subjective frequency scales for these content areas, as they did for causes of death.

Experiment 3: Direct Estimates of Event Frequencies

Experiment 1 suggested that subjects have a consistent underlying scale for the

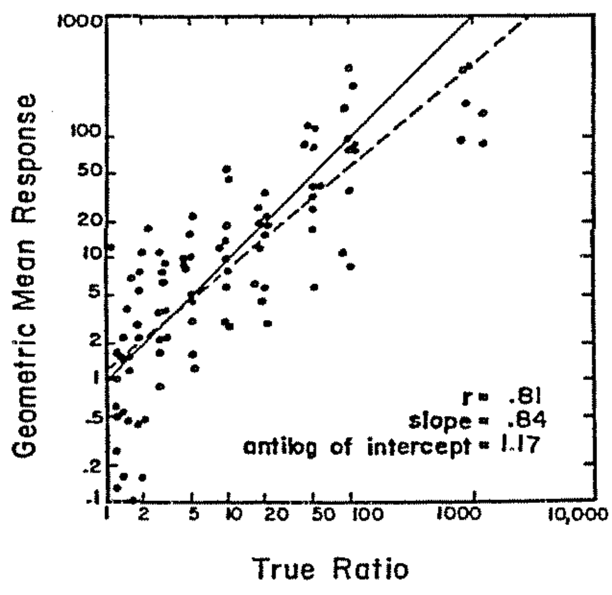

Figure 9. Geometric means of subjects' ratio judgments as a function of true ratio for 95 pairs of occupations.

frequency of lethal events, although that scale deviates markedly from the statistically correct one. Unfortunately, the incomplete paired-comparison design used in Experiment 1 did not permit the subjective scale to be uncovered for all events. When the judged relative frequencies for a given pair were in error, it was difficult to determine whether judgments were biased for one, the other, or both members of the pair. Experiment 3 elicited direct estimates to clarify the nature of the biases for individual lethal events.

Method

The subjects were 74 respondents to an advertisement in the University of Oregon campus newspaper. Each subject was assigned to one of two groups. One group of 40 subjects was told that the frequency of deaths in the U.S. due to *motor vehicle accidents* was 50,000 per year (MVA group). Using this value as a standard, they were asked to estimate the frequency for the other 40 lethal events shown in Table 1. The remaining 34 subjects (Group E) were given the standard of 1,000 deaths by *electrocution*. The glossary used in Experiment 1, which defined some of the events, was provided. The 41 events were listed in alphabetical order on a single sheet. Subjects were encouraged to erase and change answers to make the relative frequencies of the entire set consistent with their best opinions.

Since there were about 205,000,000 persons in the United States when the data were collected, the rates per 10^8 shown in Table 1 were multiplied by 2.05 to provide statistical frequencies against which to compare subjects' judgments. The standards given to the subjects, 1,000 for electrocutions and 50,000 for motor vehicle accidents, were close to these computed statistical frequencies (1,025 and 55,350, respectively).

Results

The data for one subject from Group MVA and two subjects from Group E were excluded from all analyses because they gave unreasonably high estimates (the sum of their estimates for all 41 causes of death exceeded 50,000,000, whereas the sum of the statistical frequencies is 3,553,004). Another subject was excluded from Group E because of unusually low responses. All of this subject's responses were below 1,000 (the value of the standard); 38 of 40 responses were less than 100. As a result of these exclusions, the data presented below are based on 39 subjects in Group MVA and 31 subjects in Group E.

Because arithmetic means tend to be unduly influenced by occasional extreme values, the present results are based on the geometric means of the estimates. The use of medians leads to essentially the same results. For both groups, the correlation between log geometric mean and log median was .99 (for Group MVA, slope = 1.01, antilog of intercept = .97; for Group E, slope = 1.00, antilog of intercept = 1.17).

The log geometric mean direct estimates for Groups E and MVA were highly correlated ($r = .98$). However, as shown in Table 5, the geometric means for the MVA group were larger than those for Group E for 34 of 41 causes (sign test; $p < .01$). This difference may be due to MVA subjects anchoring on a larger standard than that presented to E subjects. (The two columns in Table 5 labeled *ratio of judged to predicted* will be discussed later in the article.)

Accuracy. Figures 10 and 11 show the geometric mean judgments plotted against the true rates (excluding smallpox). The best-fitting quadratic curves are also shown. For both groups, quadratic equations provided a significantly better fit ($p < .01$) to the data than linear equations. For the MVA group, the correlation between the log geometric mean responses and the predictions from the quadratic equation was .92; the linear correlation was .89. For Group E the correlations were .93 (quadratic) and .91 (linear).

Although the log geometric mean estimates correlated highly with the true frequency, these correlations, calculated over a true frequency range of over 800,000, do not indicate substantial accuracy. Large estimation errors were evident, as with the paired-comparison judgments. For example, as Table 5 indicates, accidental death was again judged about equal in frequency to all diseases (although death from disease is 15 times more likely), cancer was judged to be about twice as frequent as heart dis-

Table 5
Results from Direct Estimates

Cause	Rate per 2.05×10^8	MVA Geometric mean	MVA Ratio of judged to predicted	Electrocution Geometric mean	Electrocution Ratio of judged to predicted
Smallpox	0	88		37	
Poisoning by vitamins	1	237	1.27	44	1.16
Botulism	2	379	1.97	88	1.96
Measles	5	331	1.39	85	1.47
Fireworks	6	331	1.54	77	1.26
Smallpox vaccination	8	38	.17	14	.22
Whooping cough	15	171	.69	51	.62
Polio	17	202	.80	47	.55
Venomous bite or sting	48	535	1.67	233	1.85
Tornado	90	688	1.82	463	2.86
Lightning	107	128	.32	64	.37
Nonvenomous animal	129	298	.71	102	.54
Flood	205	863	1.77	627	2.71
Excess cold	334	468	.81	211	.73
Syphilis	410	717	1.15	338	1.05
Pregnancy, childbirth, and abortion	451	1,932	2.98	935	2.78
Infectious hepatitis	677	907	1.19	328	.80
Appendicitis	902	880	1.03	416	.87
Electrocution	1,025	586	.65	1,000[a]	1.96
Motor-train collision	1,517	793	.74	598	.95
Asthma	1,886	769	.65	333	.47
Firearms	2,255	1,623	1.26	1,114	1.42
Poisoning	2,563	1,318	.96	778	.92
Tuberculosis	3,690	966	.59	448	.43
Fire and flames	7,380	3,814	1.62	2,918	1.86
Drowning	7,380	1,989	.85	1,425	.91
Leukemia	14,555	2,807	.81	2,220	.92
Accidental falls	17,425	2,585	.68	2,768	1.03
Homicide	18,860	8,441	2.10	3,691	1.30
Emphysema	21,730	3,009	.69	2,696	.86
Suicide	24,600	6,675	1.42	3,280	.97
Breast cancer	31,160	3,607	.66	2,436	.61
Diabetes	38,950	2,138	.34	1,019	.22
Motor vehicle accident	55,350	50,000[a]	6.34	33,884	5.76
Lung cancer	75,850	9,723	1.00	9,806	1.33
Stomach cancer	95,120	4,878	.43	2,209	.26
All accidents	112,750	86,537	6.77	91,285	9.32
Stroke	209,100	10,668	.54	4,737	.31
All cancer	328,000	47,523	1.70	43,772	2.00
Heart disease	738,000	25,900	.49	21,503	.51
All disease	1,740,450	80,779	.75	97,701	1.14

Note. MVA = motor vehicle accident.
[a] Standard.

ease (the reverse is true), floods were estimated to take more lives than asthma (asthma is 9 times more likely), diabetes was seen as only half as frequent as fire and flames, homicides were judged almost as frequent as stroke and so on.

The errors evident in the direct estimates were partitioned into primary and secondary components, as was done with the paired-comparison judgments in Experiment 1. The primary bias was an overestimation of low-frequency events and underestimation of high-frequency events by both groups. As shown by the quadratic curve in Figure

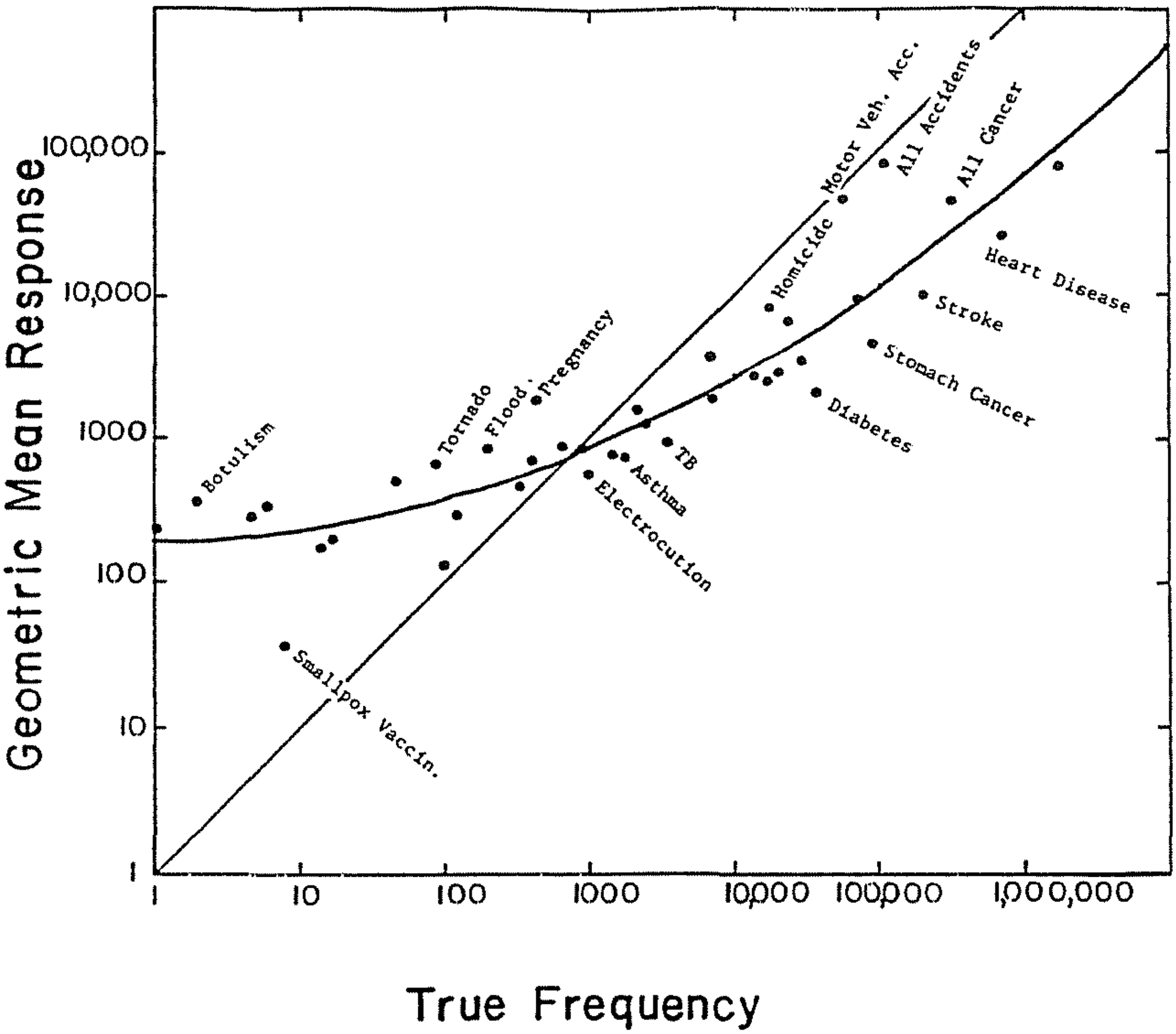

Figure 10. Geometric means (GM) of ratio judgments by motor vehicle accident group subjects as a function of true frequency (TF). (Curved line is best-fitting quadratic: $\log GM = .07 [\log TF]^2 + .03 \log TF + 2.27$.)

10, the crossover point for Group MVA was at a true rate of about 800; all events with frequencies lower than that were overestimated, and all above that point were underestimated. For Group E (see Figure 11) the crossover point was less clear; it occurred around a true rate of 250.

Secondary bias. Deviations from the regression curves were quite similar for the two groups (see Figures 10 and 11). The correlation between the two groups' residual values (i.e., the vertical distance between each point and the regression curve) was .91 across the 40 items (excluding smallpox), indicating a consistent secondary bias above and beyond the primary bias evidenced by the regression curves. The antilogs of these residuals are shown in Table 5, in the columns labeled *ratio of judged to predicted.* Some of the items with large residuals are labeled on the two figures. The similarity between the two groups of subjects, relative to their own regression lines, is striking. Frequency of death due to all accidents, motor vehicle accidents, pregnancy, flood, tornado, and cancer was relatively overestimated by both groups. Death due to smallpox vaccination, diabetes, lightning, heart disease, tuberculosis, and asthma was relatively underestimated by both.

Comparison with Experiment 1. Overall, there is a close relationship between the

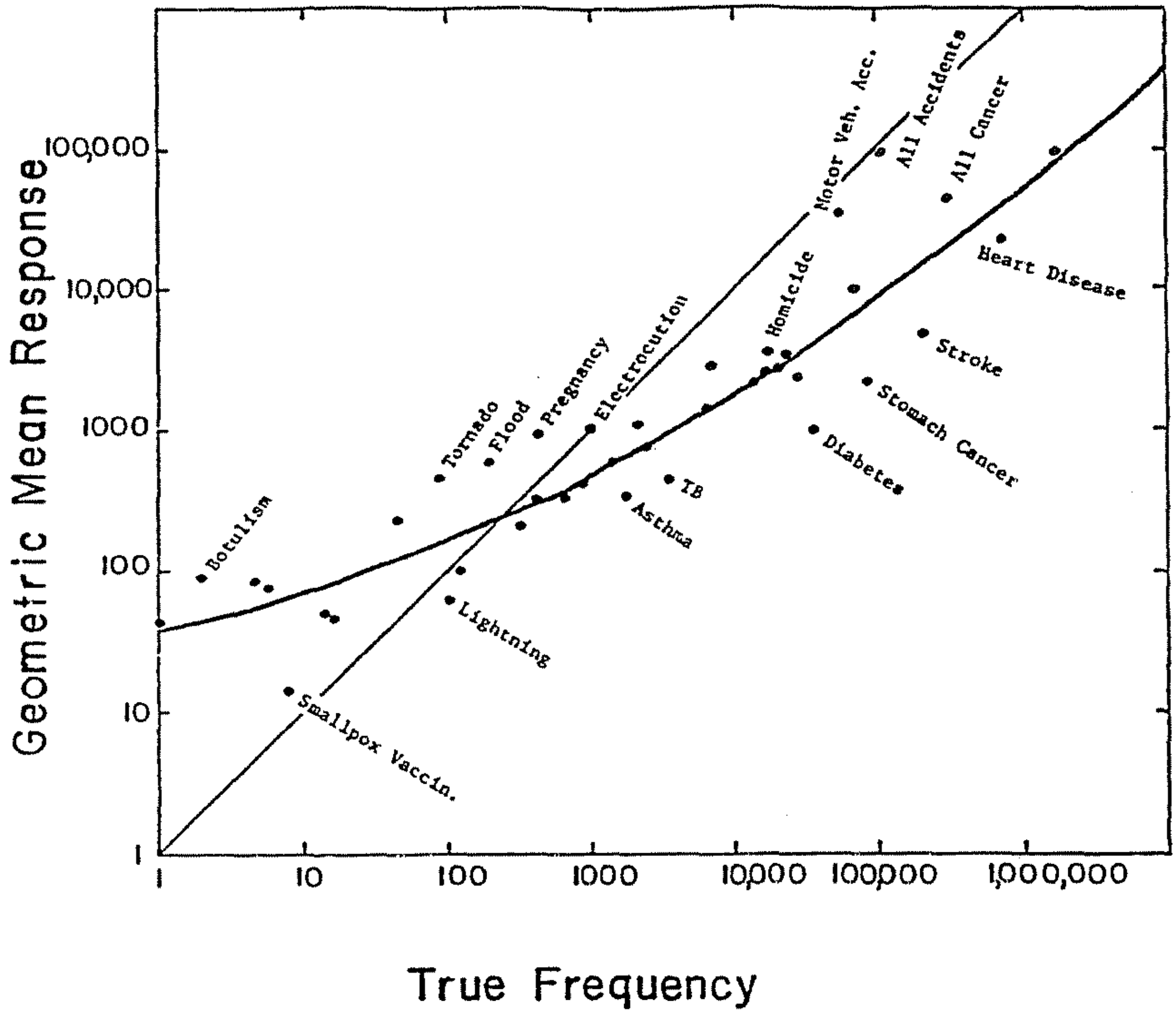

Figure 11. Geometric means (GM) of ratio judgments by electrocution group subjects as a function of true frequency (TF). (Curved line is best-fitting quadratic: log GM = .05 (log TF)^2 + .22 log TF + 1.58.)

direct estimates of the present experiment and the paired-comparison results of Experiment 1. From the geometric means of the direct estimates one can compute ratios for each of the 106 pairs studied in Experiment 1. The logs of these derived ratios were highly correlated with the logs of the geometric mean frequency ratios from Experiment 1 (college students): $r = .94$ for the MVA group and .93 for the E group (across all 106 pairs).

Neither the judged ratios from Experiment 1 nor the ratios derived from the direct estimates of the present experiment were consistently closer to the true ratios. The judged ratios from Experiment 1 were less accurate when the true ratio was low (< 10:1) and more accurate when the true ratio was high (≥ 10:1).

Individual performance. For each subject the correlation between log response and log true rate was calculated across the 40 stimuli (excluding smallpox). Individuals in Group E showed a range from .61 to .92 and a median of .77. Within Group MVA, correlations ranged from .28 to .90; the median was .66. Again, these correlations do not indicate substantial accuracy. Subjects who could make only the roughest discriminations, for example, knowing that death from botulism or lightning is less likely than death from all cancer or all acci-

Table 6
Ratings on Eight Predictor Variables

Cause	Indirect Death	Indirect Suffering	Direct Death	Direct Suffering	Newspaper frequency	Newspaper inches	Catastrophe	Conditionality
Smallpox	2.20	2.48	1.02	1.33	0	0	1.35	5.87
Poisoning by vitamins	1.23	1.43	1.00	1.07	0	0	1	4.36
Botulism	2.82	2.82	1.03	1.36	0	0	1.49	10.32
Measles	2.07	3.00	1.05	2.41	0	0	1	1.81
Fireworks	2.43	2.85	1.10	1.56	0	0	1	3.73
Smallpox vaccination	1.30	1.71	1.03	1.41	0	0	1	.71
Whooping cough	1.48	1.95	1.00	1.38	0	0	1	3.84
Polio	2.49	2.87	1.15	1.77	0	0	1	4.81
Venomous bite or sting	2.41	2.97	1.05	2.15	0	0	1	6.84
Tornado	3.46	3.75	1.07	1.38	36	153.5	4.51	6.25
Lightning	2.34	2.38	1.05	1.23	1	.8	1.01	10.06
Nonvenomous animal	2.30	2.89	1.03	1.82	4	33.8	1	3.19
Flood	3.66	4.05	1.12	1.56	4	41.8	5.57	6.52
Excess cold	2.62	2.93	1.15	1.57	0	0	1.20	10.15
Syphilis	2.51	3.67	1.07	1.79	0	0	1	5.19
Pregnancy, abortion, and childbirth	3.07	3.84	1.13	2.03	0	0	1.01	4.57
Infectious hepatitis	2.03	2.77	1.12	2.02	0	0	1	7.79
Appendicitis	2.00	2.67	1.10	2.30	0	0	1	3.53
Electrocution	2.90	2.69	1.21	1.57	5	42.2	1	15.81
Motor-train collision	3.03	2.85	1.23	1.28	0	0	2.12	14.87
Asthma	1.62	3.13	1.18	2.41	1	1.9	1	2.07
Firearm accident	3.89	3.87	1.44	1.67	8	28.2	1.02	10.34
Poisoning solid/liquid	3.02	3.05	1.10	1.61	3	17.9	1.03	10.81
Tuberculosis	2.71	3.13	1.10	1.61	0	0	1.08	7.68
Fire and flames	4.07	4.15	1.20	1.71	94	320.7	1.73	10.58
Drowning	3.82	3.23	1.69	1.68	47	247	1.07	17.65
Leukemia	3.56	3.38	1.36	1.23	1	14.8	1	15.00
Accidental falls	3.18	3.54	1.31	2.43	15	124.8	1.03	4.79
Homicide	4.69	4.33	1.39	1.23	278	5042.9	1.06	18.32
Emphysema	3.02	3.36	1.31	1.75	1	1.1	1	11.03
Suicide	4.00	3.66	1.74	1.71	29	356.7	1	17.23
Breast cancer	3.03	4.33	1.38	2.00	0	0	1	9.39
Diabetes	2.37	3.49	1.31	2.39	0	0	1	6.45
Motor vehicle accident	4.69	4.71	2.03	2.61	298	1440.5	1.64	8.97
Lung cancer	4.15	4.21	1.82	1.66	3	35.9	1	14.26
Stomach cancer	2.89	3.08	1.59	1.59	0	0	1	11.87
All accidents	4.44	4.64	2.05	2.43	715	2861.4	1.70	6.97
Stroke	3.87	3.98	1.95	2.18	12	130.7	1	11.76
All cancer	4.54	4.59	2.38	2.34	25	188.5	1	13.16
Heart disease	4.28	4.34	2.15	2.10	49	303.4	1	13.00
All disease	4.48	4.49	2.25	2.44	111	727.1	1.19	8.00
Range of scale	1–5	1–5	1–3	1–3	1–∞	1–∞	1–∞	0–20
M	3.04	3.35	1.35	1.80	42.4	295.5	1.31	8.77

dents, would show high correlations.

Experiment 4: Experience and Bias

Experiments 1 and 3 demonstrated that the frequencies of some lethal events are consistently misjudged. In hopes of learning more about the nature of these errors and biases, Experiment 4 examined people's direct and indirect experiences with these events and some of the events' special characteristics. Eight different characteristics were assessed for each lethal event and then used to predict the errors found in Experiments 1 and 3. Four of the measures assessed how much experience subjects feel they have had with the different causes of death. Two measures reflected the frequency with which causes of death appear in newspaper articles. The final measure reflected the degree to which the various causes of death were judged to be catastrophic (inflicting simultaneous multiple casualties) and lethal (inevitably producing death for people suffering from the condition).

Method

Experience ratings. A new group of 61 subjects recruited through the University of Oregon campus newspaper was asked to rate each of the 41 causes of death according to their personal experiences with the event as a cause of death and suffering.

Two ratings of *indirect* experience were obtained by asking subjects to indicate how often they had heard about the event via the media (newspapers, magazines, radio, television, etc.) as (a) a cause of death and (b) a cause of suffering (but not death). Ratings were made on a 5-point scale whose extreme categories were never (coded as 1) and often (coded as 5).

Subjects' *direct* experience with the 41 events as causes of death were elicited by having them check one of the following three statements for each event: At least one close friend or relative has died from this (Code 3); someone I know (other than a close friend or relative) has died from this (Code 2); no one I know has died from this (Code 1). Direct experience with these events as causes of suffering was elicited with similar questions, with the word *died* replaced by the phrases *suffered (but not died)*.

Thus, each subject provided four ratings for each of the 41 events. These were ratings of (a) indirect death (coded 1 to 5), (b) indirect suffering (coded 1 to 5), (c) direct death (coded 1 to 3), and (d) direct suffering (coded 1 to 3).

Newspaper coverage. The news media provide two kinds of information about causes of death. One, as noted earlier, is reports of statistical analyses (Figure 1). The other, far more prevalent, is the day-to-day reporting of fatalities as they happen. The latter is likely to be biased toward violent and catastrophic events (see, for example, Arlen's [1975] survey of television's treatment of death). Because of the potential importance of media exposure, we supplemented people's ratings of their indirect (media) experiences with a survey of newspaper reports The local daily newspaper (the *Eugene Register-Guard*) was examined on all days of alternative months for a year, starting with January 1, 1975 (for a total of 184 days). Two tallies were made for each cause of death: the total number of deaths reported and the square inches of reporting devoted to the deaths (excluding photographs).

Catastrophic ratings. Economist Theodore Bergstrom (Note 2) has asked whether catastrophic events with multiple victims in close geographic and temporal proximity will be judged as more

Table 7
Direct Estimates Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MVA LGM	—												
2. E LGM	.98	—											
3. MVA group residuals	.40	.35	—										
4. E group residuals	.36	.38	.91	—									
5. Log true frequency	.89	.91	.00	.00	—								
6. Indirect death	.85	.86	.45	.48	.74	—							
7. Indirect suffering	.86	.86	.46	.44	.76	.88	—						
8. Direct death	.90	.88	.19	.19	.82	.77	.71	—					
9. Direct suffering	.52	.50	.22	.16	.46	.21	.49	.47	—				
10. News frequency	.56	.54	.59	.56	.33	.50	.47	.45	.29	—			
11. News inches	.45	.41	.45	.36	.29	.50	.43	.31	.04	.77	—		
12. Catastrophe	−.03	.02	.29	.40	−.12	.21	.21	−.11	−.17	.09	.02	—	
13. Conditional death	.47	.51	.04	.08	.54	.65	.37	.47	−.28	.10	.30	−.07	—

Note. MVA = motor vehicle accident; E = electrocution; LGM = log geometric mean.

likely than events that take as many lives but in a less spectacular, one-at-a-time fashion. He hypothesized that catastrophes are more spectacular and thus more memorable, a speculation in keeping with availability considerations. On the other hand, the more frequent instances of noncatastrophic events may lead them to be judged more accurately, whereas casualties from catastrophic events may be underestimated because of their massed presentation (Hintzman, 1976). To assess catastrophic potential, 13 employees of the Oregon Research Institute were asked to estimate the average number of people who die from a single fatal episode of each of the 41 causes of death.

Conditional death ratings. In Experiments 1 and 3 subjects appeared to underestimate (relative to the regression line) the frequencies of deaths due to events that are common in nonfatal form, such as smallpox vaccination and asthma. One possible explanation of this error is that subjects confused $P(\text{event x}|\text{death})$ with $P(\text{death}|\text{event x})$ and failed to appreciate the importance of base rates (Tversky & Kahneman, 1974; Bar-Hillel, Note 3). Consider the question of whether a randomly selected death is more likely to be due to smallpox or smallpox vaccination. This question calls for comparing $P(\text{smallpox}|\text{death})$ with $P(\text{smallpox vaccination}|\text{death})$, the latter being statistically greater. However, subjects may be relying on $P(\text{death}|\text{smallpox})$ and $P(\text{death}|\text{smallpox vaccination})$ to answer such questions. If the base rates for the two events are discrepant (there are many more smallpox vaccinations than cases of smallpox), the resulting judgments will be in error.

To explore the role of this characteristic, 31 college students were asked to rate the probability of death given that one suffered from or experienced each condition. The ratings were made on a scale from 0 (surely won't die) to 20 (surely will die).

Results

Mean values. Mean values for the six subjective scales and the two newspaper measures are shown in Table 6. As one would expect, subjects reported greater experience with these events as causes of suffering than as causes of death. The most frequently experienced event was motor vehicle accidents, while the lowest ratings were given to poisoning by vitamins.

During 184 days of newspaper reporting, 19 of the listed causes of death were never mentioned. Some of these 19 causes are quite frequent: cancer of the digestive system, diabetes, breast cancer, and tuberculosis. In contrast, the eighth most frequently reported cause of death in the newspapers, tornadoes, is in fact relatively rare. The reported tornado deaths may represent all deaths from this cause in the United States during the dates covered. Note also that homicide, which is 23% *less* frequent than suicide, was reported 9.6 times as often, with 15 times as much space devoted to it.[5]

Few of the listed causes of death were classed as catastrophic in terms of the judged number of people dying on a single occasion. Flood, tornado, and motor vehicle/train collisions led the catastrophe ratings.

The conditional death ratings seem reasonable. The lowest rating was given to smallpox vaccination, while the highest was to homicide, followed by drowning. Some chronic diseases—asthma, diabetes, syphilis, and tuberculosis—were rated below the overall mean of 8.77, but emphysema (11.03) and heart disease (13.00) were both rated well above the mean.

Correlations: direct estimates. Correlational analyses were performed to determine whether the eight measures predict the judgments and biases found in Experiment 3. Two aspects of the direct-estimate data were predicted from the eight characteristics: (a) the log geometric mean response to the 40 lethal events (excluding smallpox) and (b) the index of secondary bias used in Experiment 3 (the signed difference between the log geometric mean of the judged frequencies and the log geometric mean predicted by the quadratic regression curves shown in Figures 9 and 10).

Table 7 shows the intercorrelation matrix for the four response variables (log geometric mean frequencies and residuals for Group MVA and for Group E), the true frequency, and the eight predictor variables. The lower left rectangle of correlations indicates the predictive power of the eight independent variables. Three of the four experience ratings showed strong correlations with the four response variables. Note that these ratings correlated more highly with

[5] This result may be even more extreme than it appears, since there is good reason to suppose that the official records we used to establish "true" rates underestimate the frequency of suicide.

the subjects' responses than with the true frequencies. The ratings of direct suffering showed only moderate correlations with subjects' responses.

News frequency and news inches were also modestly good predictors of the response variables. They were poorly correlated with true frequency, demonstrating the biased view of reality that newspapers present.[6] The catastrophe ratings showed quite low correlations with all other variables. This may be due, in part, to the lack of variance in these ratings; over half were equal to 1.0, and only 10 of 41 were greater than 1.08. Finally, conditional death ratings were moderately correlated with the geometric mean responses, but not with the residuals.

The correlations among the eight predictor measures are also shown in Table 7. Indirect death, indirect suffering, and direct death ratings showed fairly high intercorrelations but lower correlations with direct suffering. The two newspaper measures were highly intercorrelated. However, these newspaper measures correlated only moderately with the indirect death ratings, even though the instructions for the latter task emphasized newspaper coverage.

The direct estimates made by the subjects in Experiment 3 may have been biased because they were influenced by past experience with *indirect* sources of information (such as newspapers), which themselves were biased. We suspected that ratings of direct experience might be less biased and, therefore, might provide more accurate estimates of the true frequencies than did the direct estimates of frequency. This hypothesis was tested and was not supported. Although the direct death ratings did correlate more highly with the true frequency ($r = .82$) than did any of the other predictor measures, the direct estimates of Experiment 3 did even better ($r = .89$ and .91).

Correlations: paired comparisons. Similar correlational analyses were performed relating the eight measures with the paired-comparison judgments of Experiment 1. To do this, a difference score was formed on each measure for each of the 101 pairs (excluding smallpox) by subtracting the score associated with the less likely cause of death from the score associated with the more likely cause of death. These difference scores were then correlated with four dependent variables (the log geometric mean responses and the index of secondary bias used in Experiment 1, for students and for league members), with the log true ratio, and with each other. The resulting correlation matrix is not shown here, because it was quite similar to Table 7.

As with the direct-estimate data, the ratio of the direct death ratings correlated with true ratio more highly ($r = .62$) than did any of the other predictor measures. However, it could not successfully be substituted for the judged ratios of Experiment 1 in an attempt to improve accuracy, since the judged ratios were even more highly correlated with true ratio ($r = .69$ for students and .75 for league members).

Regression analyses predicting responses and biases. To bring greater clarity to this mass of correlations, eight stepwise regressions were performed. Four of these analyses predicted the log geometric mean responses of the four separate groups of subjects: students' paired comparisons, league members' paired comparisons, Group E's direct estimates, and Group MVA's direct estimates. The other four stepwise regression analyses predicted secondary bias (the residuals from the correlations of each of these four groups with the statistical frequencies).

The predictor variables for each of the stepwise regressions were the eight measures previously described, using differences between 101 pairs to predict the paired-comparison data or 40 mean ratings to predict the direct estimates and their residuals.

Because of the instability of stepwise regression solutions with highly intercorrelated predictors, our primary criterion for variable selection was replicability. Only variables that entered the equations for both league and student subjects in Experiment

[6] Similar evidence of bias in another newspaper may be found in Combs and Slovic (Note 4).

Table 8
Variables Emerging from Stepwise Multiple Regressions in Both Replications

Dependent variables			
Log geometric mean		Residuals	
Paired comparisons	Direct estimates	Paired comparisons	Direct estimates
Indirect suffering Direct death	Indirect suffering Direct death News frequency	Indirect death Direct death Conditional death[a]	News frequency catastrophe

[a] Negative weight.

1 or both Group E and Group MVA in Experiment 2 are discussed. Table 8 lists the variables that emerged from *both* groups of subjects. The inclusion criterion was an *F* to enter[7] of 3.0 or greater. The log geometric means were highly predictable, with *R*s ranging from .88 to .96 using just three of the eight predictors. The residuals were also predictable, with *R*s ranging from .64 to .80 using the variables selected by the stepwise regression.

Two variables, indirect suffering and direct death, did most of the job of predicting the subjects' log geometric mean responses for both paired comparisons and direct estimates. The regressions on the residuals show a more mixed pattern. For the residuals from the paired-comparison data, three predictors were common to both the student and league data: indirect death, direct death, and conditional death. Conditional death had a negative weight because of its low correlation with the dependent variable and its high correlation with indirect death. For the prediction of residuals from the direct estimates, news frequency and catastrophe ratings were the only predictors that were significant in both groups. In view of the highly skewed distributions of these two measures, it is somewhat surprising to see them emerge as valid predictors. However, news frequency correlated with direct-estimate residuals higher than did any other single predictor. Of the 7 causes of death with catastrophe ratings of 1.5 or greater, six (all accidents, motor vehicle accidents, flood, botulism, tornado, and fire and flames) were among the 10 causes of death with the highest residuals (i.e., the 10 most overestimated causes of death, relative to the regression line).

The above analyses indicate that measures tapping the availability of information about causes of death do a good job of predicting subjects' judgments of the frequencies and relative frequencies of these causes of death. Further, we have shown that the consistent errors people make (the secondary bias) can be predicted from salient features of the events such as their catastrophic nature and from ratings of experience with the lethal events made by a different group of subjects.

Experiment 5: Debiasing

Experiments 1 and 3 showed that subjects make severe and consistent errors in judging the frequency or relative frequency of lethal events. Experiment 5 was designed to see if subjects could correct these errors when they were told the hypothesized causes of the errors. Emphasis was placed on the secondary bias and its possible causes: uneven newspaper coverage and the effects of imaginability and memorability.

Study 5A

Method

In Study 5A, subjects made paired comparisons for 31 of the 106 pairs of Experiment 1. Twenty-one of these pairs were severely misjudged in Experiment 1 (either the percentage correct was

[7] An "*F* to enter" tests the significance of the increase in the proportion of explained variance achieved by including an additional variable in the regression equation.

less than 60% or the geometric mean was off by a factor of 9 or more). The geometric means of the remaining 10 were estimated moderately well (within a factor of 1.5). The present study was conducted with a college student population similar to that in Experiment 1 and with the same instructions except that one group, the *debiasing* group ($n = 30$), was given the following special information:

> Note: In a previous study of this kind we found that, for some pairs, the relative likelihoods were greatly misperceived. Sometimes the ratio of the more likely to the less likely item was judged to be much greater than it really was. In other cases the ratio was judged much too small or even in the wrong direction; that is, the less likely item was judged to be more likely.
>
> We believe that when people estimate these likelihoods, they do so on the basis of a) how easy it is to *imagine* someone dying from such a cause, b) how many instances of such an event they can *remember* happening to someone they know, c) *publicity* about such events in the news media, or d) *special features* of the event that make it stand out in one's mind.
>
> Reliance on imaginability, memorability, and media publicity, although often useful, can lead to large errors in judgment. When events are disproportionately imaginable or memorable, they are likely to be overestimated. When they are rather unmemorable or unpublicized or otherwise undistinguished, they are likely to be underestimated. Events such as ulcers that are common, but usually non-fatal, may also be underestimated because people tend to imagine or remember them in their non-fatal form.
>
> Try not to let your own judgments be biased by factors such as imaginability, memorability, or media publicity.

A control group ($n = 22$) also judged the 31 pairs without receiving any special instructions.

Results

Examination of percentage correct revealed no evidence for debiasing. The original subjects (Experiment 1) were best on 9 pairs, the control subjects were best on 12 pairs, and the debiasing group subjects were best on 10 pairs.

A further search for improvement in the data of Study 5A can be made by comparing the ratio judgments of these two new groups of subjects either with the true ratios (under the assumption that the instructions exhorted the subjects to come closer to the truth) or with the ratios predicted from the regression analysis of the original subjects (under the assumption that the instructions emphasized the nature of the secondary bias, not the primary bias). No evidence for effective debiasing can be seen under either comparison. For geometric means, when the comparison is made to the true ratio, the original group was best on 12 pairs, the controls on 6 pairs, and the debiasing group on 13 pairs. When compared with the predicted ratios, the original group was best on 12 pairs, the control group on 7, and the debiasing group on 12. Looking only at the 21 pairs that were originally judged poorly, there is still no evidence of improvement in the debiasing group. Even those pairs on which the debiasing group did best showed only modest improvement. For example, death by diabetes is 95 times more likely than death by syphilis. The debiasing group was "superior" in giving a geometric mean response of 9.7 rather than the original group's geometric mean of 2.4. Death by stroke is 102,000 times more likely than death by botulism. The value predicted by the regression analysis of the original subjects was 1,002. Those original subjects showed a strong secondary bias; their geometric mean response was 106. The debiasing group gave a mean response of 135.

Study 5B

Method

A second debiasing study was undertaken to provide subjects even more opportunity for using knowledge of the secondary biases to improve their performance. The subjects, drawn from the same student population, were shown 19 pairs of events. The instructions indicated that each of these pairs had been seriously misjudged in an earlier experiment (which was the case). For each pair, the subjects were given the response from Experiment 1 and were asked to improve it, that is, to give a new response that they thought would be closer to the true ratio.

The instructions for a debiasing group of 29 subjects included a discussion of the presumed sources of error, illustrated with several examples showing the possible effects of personal experience, media publicity, imaginability, and the like on previous subjects' judgments. A control group of 27 subjects did not receive this additional discussion.

The instructions read as follows. Brackets indicate material shown only to the debiasing group.

> We recently studied the ability of University of Oregon students to judge the likelihood of various causes of death in the United States.
>
> For example, subjects were given a pair of events such as:
> A. Measles,
> B. Tornado.
>
> They were asked: Which causes more deaths annually in the U.S., A or B? They were also asked to estimate how many times more likely the more frequent cause of death was compared to the less frequent of the two.
>
> We found that, for some pairs, the relative likelihoods were greatly misjudged. Sometimes the ratio of the more likely to the less likely item was judged much too small or even in the wrong direction; that is, the less likely item was judged to be more likely.
>
> [We believe that when people estimate these frequencies, they do so on the basis of a) how easy it is to *imagine* someone dying from such a cause, b) how many instances of such an event they can *remember* happening to someone they know, c) *publicity* about such events in the news media, or d) *special features* of the event that make it stand out in one's mind.]
>
> [When events are disproportionately imaginable or memorable, they are likely to be overestimated. When they are rather unmemorable or unpublicized or otherwise undistinguished, they are likely to be underestimated. Events such as accidental falls, that are common but usually non-fatal, may also be underestimated because people tend to imagine or remember them in their non-fatal form.]
>
> On the following pages there are 19 pairings of death-producing events. The relative likelihood of the more common to the less common event was greatly misperceived in each of these pairs.
>
> [We want to see whether you can reduce the magnitude of the errors for these pairs. To do this think about how factors such as media coverage or ease of imagining or remembering the event as a cause of death are likely to work to bias the judgments for each of the pairs.]
>
> Here are some examples to illustrate the task:
>
> | | Previous Answer | Your Answer |
> |---|---|---|
> | A. Hepatitis | B 4.55 | |
> | B. Drowning | | |
>
> The average subject chose B as more likely and judged it to be 4.55 times more likely than A. Which would you choose and what ratio would you give?
>
> Actually, the correct answer is B and the true ratio is 10.9 to 1. We see that the average subject overestimated Hepatitis relative to Drowning. [Maybe this is because of the special attention given by the media to Hepatitis, especially in relation to abuse of hypodermic needles.]
>
> Try this one:
>
> | | Previous Answer | Your Answer |
> |---|---|---|
> | A. Leukemia | A 1.30 | |
> | B. Accidental Falls | | |
>
> The average subject thought death from leukemia was 30% more common (ratio 1.30 to 1) than death from falls. However, death from falls is really 20% more frequent. So the correct answer is B with a ratio of 1.20. [The error may stem from the dramatic nature of leukemia and the greater amount of media publicity it receives, or it may stem from the fact that accidental falls are common but usually non-fatal.]
>
> For a final example, consider:
>
> | | Previous Answer | Your Answer |
> |---|---|---|
> | A. Poisoning by solid or liquid | A 5.26 | |
> | B. Tuberculosis | | |
>
> The average subject thought death by poisoning was 5.26 times more likely than death from tuberculosis. However, death from tuberculosis is really 44% more frequent than death from poisoning so the correct answer is B with a ratio of 1.44. [Again, it is easy to see how media publicity regarding poisoning and the dramatic nature of the event could cause subjects to overestimate it compared to the drab, undramatic, perhaps old-fashioned disease, tuberculosis.]
>
> Note that a ratio of 1.20 means 20% more likely, 1.50 means 50% more likely, 1.80 means 80% more likely, etc.
>
> For each pair, write the letter of the item you think is a more likely cause of death and give your judgment about how many times more frequent the more frequent item is.

Results

The special instructions given to the debiasing group had no effect on performance. Neither the debiasing group nor the control group was able to improve consistently upon the mean responses given by subjects in Experiment 1. For each pair, we calculated the percentage of subjects in the debiasing group and in the control group whose responses were closer to the true ratio than was the geometric mean of the original Experiment 1 group. We also calculated the percentage of subjects in both groups whose responses were closer to the ratio predicted from the Experiment 1 regression line (i.e., who had smaller secondary bias). In every case the percentage closer to the true ratio was equal to the percentage closer to the regression line. The average percentage of improved answers was only 53.8% for the experimental group (range of 21%–82%) and 52.4% for the control group (range of 37%–70%). The experimental group showed a better improvement percentage than the control group on 10 pairs, the control group was better for 8 pairs, and there was a tie on 1 pair.

Discussion

Psychological Significance

As in laboratory studies, our subjects exhibited some competence in judging frequency. Frequency estimates for causes of death, words, and occupations generally increased with increases in true frequency; similarly, the discriminability of causes increased with the ratio of their statistical frequencies. Furthermore, our subjects' assessments of the frequencies of causes of death, both direct estimates and paired comparisons, correlated more highly with the true answers than did any other measures, such as newspaper reportage and ratings of direct experience with the causes of death.

Despite the sensitivity of judgments to true frequency, the overall accuracy of both paired comparisons and direct estimates of frequency was quite poor. Unless the true frequencies of a pair of lethal events differed by more than a factor of two, there was no guarantee that subjects could correctly indicate which was more frequent. Large errors were present in the judged ratios for many pairs of events. The high correlations between direct estimates and true frequency across almost a million-to-one range of the latter variable are deceptive. Large errors were present in these estimates, much as with the paired-comparison judgments.

Primary bias. Experiments 1 and 3 demonstrated a strong primary bias, consisting of overestimation of low frequencies and underestimation of both high frequencies and large ratios, much as has been found before by Attneave (1953), Teigen (1973), and others (Poulton, 1973). We considered and rejected two possible reasons for this primary bias. One is that subjects avoid using extremely high (or low) numbers in making their responses. The absence of such biases with the words and occupations tasks of Experiment 2 makes this hypothesis implausible. Second, the underestimation of high ratios in Experiment 1 was not simply an artifact of averaging correct and incorrect answers. This is shown by the persistence of the effect for pairs in which nearly everyone got the correct answer.

Another possible explanation of the primary bias is that it results from anchoring: Subjects first choose some representative value and then adjust upward or downward according to whatever considerations seem relevant to the case at hand. Studies of anchoring and adjustment procedures have shown that such adjustments tend to be insufficient (Lichtenstein & Slovic, 1971; Tversky & Kahneman, 1974). A number of laboratory studies of frequency estimation can be interpreted as showing a tendency to anchor on the average frequency in the lists learned (see Rowe & Rose, 1977). Insufficient adjustment would produce too flat a curve, a finding often noted in laboratory studies (see Hintzman, 1976). Perhaps the clearest evidence of anchoring may be found in Experiment 3, in which the one true frequency given to the subjects could easily have served as an anchor value. Group MVA, who were given a high anchor (50,000), generally assigned higher values to the items than did Group E, whose anchor value was 1,000.

In the paired-comparison tasks no such clear-cut anchor was provided. Nonetheless, Poulton (1968) has shown that in magnitude-estimation studies, the subjective magnitude of the first stimulus presented serves as an anchor for subsequent judgments. This view is supported by Carroll's (1971) finding of a .66 correlation between the log of individual subjects' first estimates and the mean log of all their responses in estimating word frequency. The present paired-comparison data are consistent with the notion that the response to the first stimulus serves as an anchor. The causes-of-death groups received, as their first stimulus, a pair they judged as having a relatively low ratio (pair 40; geometric mean response = 4.3 for students and 18.0 for league members), while the words and occupations groups' first stimulus was judged with relatively high ratios (116 and 265, respectively). Both causes-of-death groups showed more underestimation of high ratios than did the words and occupations groups, as Poulton would predict.

Yet another possible explanation of the primary bias derives from the availability heuristic (Tversky & Kahneman, 1973), which states that assessments of frequency or probability are based on the number of instances of the event that come to mind. Cohen (1966) has found that when subjects manage to recall any of the words in a category the mean number of words recalled per category is relatively independent of the number of words in that category. If this tendency is true also for categories learned outside the laboratory, such as causes of death, and if, as suggested by Tversky and Kahneman, people base their assessments on these all-too-equal recollections, a flattening of their responses, as observed, would result.

Secondary bias. Subjects' responses exhibited numerous strong and consistent secondary biases. Some portion of these errors may be due to the unrepresentative coverage of these causes of death in the news media. Others have also speculated about the effects of such media bias. For example, Zebroski (Note 5) blamed the media for people's concerns about nuclear reactor safety. He noted that "fear sells"; the media dwell on potential catastrophes and not on the successful day-to-day operations of power plants. Author Richard Bach (1973) made a similar observation about the fear shown by a young couple going for their first airplane ride:

> In all that wind and engineblast and earth tilting and going small below us, I watched my Wisconsin lad and his girl, to see them change. Despite their laughter, they had been afraid of the airplane. Their only knowledge of flight came from newspaper headlines, a knowledge of collisions and crashes and fatalities. They had never read a single report of a little airplane taking off, flying through the air and landing again safely. They could only believe that this must be possible, in spite of all the newspapers, and on that belief they staked their three dollars and their lives. (p. 37)

The present results suggest that the media have important effects on our judgments, not only because of what they don't report (successful plane trips or reactor operations), but because of what they do report to a disproportionate extent.

Subjects may also be misinformed because of bias in their direct exposure to the various causes of death. Young people, such as our student subjects, may be underexposed to death from diseases associated with age, such as stroke, stomach cancer, and diabetes, all of which were underestimated, and overexposed to death from motor vehicle accidents, all accidents, and pregnancy, all of which were overestimated relative to the regression line.

The two explanations of secondary bias given above assume that the bias occurs because the information received by the subject is inadequate or misleading. Another explanation can be found by examining hypotheses about the biases induced by people's cognitive storage and retrieval processes. Tversky and Kahneman's (1973) concept of availability, with its emphasis on vivid or sensational events, seems relevant. Examination of Figures 9 and 10 shows that among the most overestimated causes of death (relative to the regression line) were botulism, tornado, flood, homicide, motor vehicle accidents, all accidents, and cancer. These are all sensational events. Most of the causes of death that were underestimated (relative to the regression line)—asthma, tuberculosis, diabetes, stomach cancer, stroke, and heart disease—seem to be undramatic, quiet killers.

Some of the evidence of secondary bias is inconsistent with previous laboratory findings. One such finding is that more concrete and imaginable words are judged to be *less* likely than equally frequent abstract words (e.g., Ghatala & Levin, 1976). While we had no direct measure of imaginability, one might assume that catastrophic events and those more heavily reported in the media tend to be more concrete and imaginable. However, all three of these surrogate measures of imaginability (catastrophe, news frequency, and news inches) were positively correlated with the residuals (for both paired comparisons and direct estimates). Thus, in this sense imaginable events tended to be judged *more* likely, as predicted by availability considerations.

Another difference between the present research and previous studies is found with catastrophic causes of death whose occurrences tend to be massed rather than distributed over time. Laboratory studies (e.g., Rowe & Rose, 1977) have consistently found that massing the occurrences of a word in a learned list tends to decrease its estimated frequency. Two explanations offered for this effect (Hintzman, 1976) are (a) encoding variability—spaced repetitions are more likely to receive differential coding than massed items—and (b) deficient processing of massed items. In the current experiments, catastrophic (massed) events tended to be overestimated relative to the regression line. One key difference between the usual laboratory experiments and the present study is that the former do not use stimuli that become sensational or emotionally charged when massed. Such special characteristics may lead to extra processing, rather than to deficient processing, for catastrophic causes of death.

When we have been able to compare the present results with previous laboratory work, we have found about as many mismatches as matches. The present study is based on material our subjects have learned in the real world; in most other laboratory work, the subjects were tested on material they had learned in the laboratory. Mandler (Note 6) has speculated on this difference:

> In terms of *presentation* of to-be-remembered material, the laboratory experiment fails—in comparison with the real world—with respect to three major problems: Frequency, salience, and context. The laboratory experiments fail with respect to frequency because the typical event that an individual must recall or recognize in everyday life has been encountered anywhere from a few to thousands of times; in the laboratory we look at the few and rarely look at the thousands. Salience must be of interest because encoding operations in the real world typically take place with particular attention to the relevance or salience of a particular event to other aspects of the mental apparatus; we encode what is important, while in the laboratory we are required to encode what is unimportant. Furthermore, the context of real world memory involves not simply a restricted number of materials presented in the laboratory, together with a computer or a memory drum, but rather the larger context of the individual's current plans and intentions, geographic location, and social conditions. (pp. 3–4)

Improving Judgments

One question raised by this study is how to improve intuitive judgments of frequency. We did not attempt here to correct the primary (overestimation/underestimation) bias. Work by Teigen (1973) suggests that this can be done by asking people to allocate frequencies as percentages of the total rather than having them estimate absolute numbers. This technique, however, might not prove helpful when (as with causes of death) the largest frequency is over a million times larger than the smallest frequency. It would be exceedingly difficult for subjects to express ratios even as high as 3,000:1 (as they did in the present study) using a percentage response mode. Statistical correction, using regression equations, might be the best way to correct the primary bias.

Since the secondary bias observed here seems linked to availability, we hoped to reduce that bias by informing subjects about its probable source. This information was not useful. The failure of such frontal attacks to eliminate biases (see also Fischhoff, 1977) suggests some directed restructuring of judgment tasks may be necessary. For example, Selvidge (1972) proposed having people make probability and frequency judgments on a scale in which other familiar events serve as marker points. In composing such a scale, great care would have to be taken to use only events whose subjective ordering fits their true ordering. Beyth-Marom and Fischhoff (1977) have shown that requiring people to work hard to produce specific examples of classes of events before estimating the frequencies of the classes can partially reduce availability bias. Another promising suggestion comes from Armstrong, Denniston, and Gordon (1975), who found that numerical estimates can be improved by having estimators decompose the original question into a series of subquestions about which they are more knowledgeable and whose answers lead logically to the estimate of interest. For example, an answer to the question "How many people were killed in motor vehicle accidents in the United States in 1970?" might be improved by having people answer the following related questions: (a) What is the population of the U.S.? (b) How many automobile trips does the average U.S. citizen take in a year? (c) What is the probability of a fatal injury on any particular trip? From the answers to these questions, one can calculate an answer to the original question.

Societal Implications

Economist Frank Knight once observed that "We are so built that what seems reasonable to us is likely to be confirmed by experience or we could not live in the world at all" (Knight, 1921, p. 227). But the present study and a growing body of other research (e.g., Kunreuther et al., 1978; Slovic, Kunreuther, & White, 1974; Kates, Note 1) indicate that in the evaluation of risks and hazards, Knight's optimistic assessment of human capabilities is wrong. People do not have accurate knowledge of the risks they face. As our society puts more and more effort into the regulation and control of these risks (banning cyclamates in food, lowering highway speed limits, paying for emergency coronary-care equipment, etc.), it becomes increasingly important that these biases be recognized and, if possible, corrected. Improved public education is needed before we can expect the citizenry to make reasonable public-policy decisions about societal risks (Slovic, Fischhoff, & Lichtenstein, 1976; Slovic et al., 1974). And the experts who guide and influence these policies should be aware that when they rely on their own experience, memory, and common sense, they, too, may be susceptible to bias.

We have, by necessity, studied sources of judgmental error in situations for which good estimates of true frequency exist. But our society must often make judgments about hazardous activities for which adequate statistical data is lacking, such as recombinant DNA research or nuclear waste disposal. We suspect that the biases found here (overestimation of rare events, underestimation of likely events, and an undue influence of drama or vividness) may be operating, indeed, may even be amplified, in such situations.

Reference Notes

1. Kates, R. W. *Hazard and choice perceptions in flood plain management* (Research Paper No. 78). Chicago: University of Chicago, Department of Geography, 1962.
2. Bergstrom, T. C. Preference and choice in matters of life and death. In J. Hirshleifer, T. Bergstrom, & E. Rappoport, *Applying cost-benefit concepts to projects which alter human mortality* (Rep. UCLA-ENG 7478). Los Angeles: University of California, School of Engineering and Applied Science, 1974.
3. Bar-Hillel, M. *The base-rate fallacy in probability judgments* (Rep. 77-4). Eugene, Oreg.: Decision Research, 1977.
4. Combs, B., & Slovic, P. *Causes of death: Biased newspaper coverage and biased judgments.* Unpublished manuscript, 1978. (Available from Decision Research, 1201 Oak St., Eugene, Oregon 97401.)
5. Zebroski, E. L. Attainment of balance in risk-benefit perceptions. In D. Okrent (Ed.), *Risk-benefit methodolgy and application: Some papers presented at the Engineering Foundation Workshop, Asilomar,* Calif. (Rep. UCLA-ENG 7598). Los Angeles: University of California, School of Engineering and Applied Science, 1974.
6. Mandler, G. *Memory research reconsidered: A critical review of traditional methods and distinctions* (Rep. No. 64). San Diego: University

of California, Center for Human Information Processing, 1976.

References

Arlen, M. J. The cold, bright charms of immortality. *The New Yorker*, January 27, 1975, pp. 73–78.

Armstrong, J. S., Denniston, W. B., & Gordon, M. M. The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance*, 1975, *14*, 257–263.

Attneave, F. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 1953, *46*, 81–86.

Bach, R. Nothing by chance. *The American Way*, 1973, *6*, 32–38.

Beyth-Marom, R., & Fischhoff, B. Direct measures of availability and frequency judgments. *Bulletin of the Psychonomic Society*, 1977, *9*, 236–238.

Burton, I., Kates, R. W., & White, G. F. *The environment as hazard*. New York: Oxford University Press, 1978.

Carroll, J. B. Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior*. 1971, *10*, 722–729.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 182–187.

Coombs, C. H., Dawes, R. M., & Tversky, A. *Mathematical psychology: An elementary introduction*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Estes, W. K. The cognitive side of probability learning. *Psychological Review*, 1976, *83*, 37–64.

Fischhoff, B. Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 349–358.

Ghatala, E. S., & Levin, J. R. Phenomenal background frequency and the concreteness/imagery effect in verbal discrimination learning. *Memory & Cognition*, 1976, *4*, 302–306.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, *80*, 139–145.

Hintzman, D. L. Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press, 1976.

Hintzman, D. L. *The psychology of learning and memory*. San Francisco: Freeman, 1977.

Howell, W. C. Representation of frequency in memory. *Psychological Bulletin*, 1973, *80*, 44–53.

Kates, R. W. *Risk assessment of environmental hazard*. Chichester, England: Wiley, 1978.

Knight, F. H. *Risk, uncertainty, and profit*. New York: Houghton-Mifflin, 1921.

Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence. R.I.: Brown University Press, 1967.

Kunreuther, H., Ginsberg, R., Miller, L., Sagi, P., Slovic, P., Borkin, B., & Katz, N. *Disaster insurance protection: Public policy lessons*. New York: Wiley, 1978.

Lichtenstein, S., & Slovic, P. Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 1971, *89*, 46–55.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, *68*, 29–46.

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, *69*, 1–19.

Poulton, E. C. Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, 1973, *80*, 113–121.

Rowe, E. J., & Rose, R. J. Effects of orienting task, spacing of repetitions, and list context on judgments of frequency. *Memory & Cognition*, 1977, *5*, 505–512.

Selvidge, J. *Assigning probabilities to rare events*. Unpublished doctoral dissertation, Harvard University, 1972.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976.

Slovic, P., Kunreuther, H., & White, G. Decision processes, rationality and adjustment to natural hazards. In G. F. White (Ed.), *Natural hazards: Local, national and global*. New York: Oxford University Press, 1974.

Teigen, K. H. Number and percentage estimates in sequential tasks. *Perceptual and Motor Skills*, 1973, *36*, 1035–1038.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, *5*, 207–232.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, *76*, 559–573.

U.S. Bureau of the Census. Census of population: 1970. Occupation by industry. *Final Report PC(2)-7C*. Washington, D.C.: U.S. Government Printing Office, 1972.

White, G. F. (Ed.). *Natural hazards: Local, national and global*. New York: Oxford University Press, 1974.

Received January 17, 1978
Revision received June 23, 1978 ■