APPLIED COGNITIVE PSYCHOLOGY, VOL. 11, 495–506 (1997)

When a 12.86% Mortality is More Dangerous than 24.14%: Implications for Risk Communication

KIMIHIKO YAMAGISHI*
Shukutoku University, Saitama, Japan

SUMMARY

Participants assessed the riskiness of 11 well-known causes of death. Each participant was presented with an estimation of the number of deaths in the population due to that particular cause. The estimates were obtained from a previous study of naive participants' intuitive estimations. For instance, based on the result of the previous study, the number of deaths due to cancer was presented as: '2,414 out of 10,000', '1,286 out of 10,000', '24.14 out of 100' and '12.86 out of 100'. The estimates of deaths were presented in analogous ways for the remaining ten causes of death. It was hypothesized that the judged degree of riskiness is affected by the number of deaths, irrespective of the total possible number (such as 10,000 or 100). Results from Experiment 1 were consistent with this prediction. Participants rated cancer as riskier when it was described as 'kills 1,286 out of 10,000 people' than as 'kills 24.14 out of 100 people', and similar results were observed regarding the remaining 10 causes of death. Experiment 2 replicated this trend. Implications for risk communications are discussed. © 1997 John Wiley & Sons, Ltd.

Appl. Cognit. Psychol. **11**: 495–506 (1997)
No. of Figures: 2 No. of Tables: 3 No. of References: 23

Risk research is increasingly relevant in today's society. The modern lifestyle creates unnatural hazards, such as exposure to nuclear waste or asbestos, whose potential harm is not always clear. Therefore, such harm needs to be assessed, and misassessment of risk may be costly. Slovic, Layman, Kraus, Flynn, Chalmers, and Gesell (1991) investigated Nevada residents' reactions to a proposed nuclear waste repository. They claimed that the intensely negative image of possible risks may cause people to overlook economic prospects, and hence cause loss to the State. New York City spent $83 million for removal of asbestos from its public schools in the fall of 1993 ('The asbestos panic attack', 1995). Yet, the lifetime probabilities of premature death due to asbestos in school buildings and being a pedestrian hit by a car are 1/100,000 and 1/290, respectively. One might argue that the tax money could

---

Preliminary reports of this work were presented at the annual meetings of the Society for Judgment and Decision Making, St. Louis, U.S.A., November 1994, and the Japanese Psychological Association, Tokyo, Japan, September 1996. I would like to thank Richard Gonzalez, John Miyamoto, Karl Halvor Teigen, and participants of the Judgment and Decision Making seminar at the University of Washington for helpful suggestions.

*Correspondence to: Kimihiko Yamagishi, School of Business and Environment, Shukutoku University, Fujikubo 1150-1, Miyoshi, Iruma, Saitama 354, Japan. Address e-mail to kimihiko@shukutoku.ac.jp.

CCC 0888–4080/97/060495–12$17.50

Received 7 August 1996
Accepted 22 January 1997

have been better spent if it had been allocated to a safe-driving campaign. These examples illustrate why it is important to make proper risk assessments. This paper addresses how such misassessments may occur, and in what sense such assessments could meaningfully reflect subjective perception of risk.

Research on risk perception has shown two aspects of intuitive judgments on environmental risks. Assessment of risk has exhibited both consistencies with real-world occurrence of risks, and systematic biases. Fischhoff and MacGregor (1983) observed that subjective estimation of life-threatening events was positively correlated with the actual frequency of deaths, yet higher-frequency risks tended to be underestimated whereas lower-frequency risks tended to be overestimated (see also Fischhoff, Slovic and Lichtenstein, 1981). Yamagishi (1994a,b) compared frequency estimates of deaths due to 11 well-known causes as a function of different methods of judgment elicitation. When a 'narrow response range' was used, participants estimated the number of deaths in terms of, for instance, 'Out of 100 people, how many die of cancer?' When a 'wide response range' was used, the same frequency was estimated 'Out of 10,000 people'. The results showed systematic response-range effects, wherein the estimates were proportionally greater in the narrow (cancer example: 24.14%) than in the wide (12.86%) range. None the less, such estimates in both ranges were highly correlated with the actual frequencies of occurrence.

This paper aims at extending the line of research of Yamagishi (1994a,b) by asking the following question: would the estimated frequencies relative to different response ranges represent comparable seriousness of risk? Would a '1,286 out of 10,000' mortality appear as risky as a '24.14%' mortality? Logically speaking, both numbers summarize responses to formally equivalent questions. Hence the mean judged frequencies (such as 1,286 or 24.14) are 'matched' responses regarding the particular methods of judgment elicitation (10,000 or 100). In this sense, these matched responses should express equivalent degrees of riskiness.

However, it has been observed that people respond differently to equivalent forms of relative frequency information that are presented in superficially different ways. McFarland and Miller (1994) administered a test on 'social perceptiveness ability'. Subsequently, participants were provided with fictitious feedback that their performance was ranked in the 30th percentile. Participants in one group were told that they ranked third among ten people who took the test. Another group was told that they ranked 300th among 1,000 people. This manipulation of group size changed participants' self-assessment of the ability to accurately interpret social situations. As the group size increased, pessimistic participants reported lower ability levels, whereas optimistic participants reported higher ability levels. In short, in interpreting information that represents relative frequencies, people tend to be more sensitive to rote frequencies relative to larger total frequencies (such as 300 out of 1,000 rather than 3 in 10) than percentages and percentiles. Denes-Raj and Epstein (1994) presented participants with a pairwise choice of gambles. Option 1 offered a 1-out-of-10 chance (10%) of winning $1. Option 2 offered a 9-out-of-100 chance (9%) of winning the same amount. The participants tended to choose Option 2 in favour of rote frequency over percentage. When they were asked to justify their choice, they admitted that this choice went contrary to what a rational individual should do, namely to choose the 10% option.

Likewise, classic research on contingency judgment also reports how percentage judgments may be misled by rote frequencies. Smedslund (1963) presented 100 clinical cases to a group of nurses. Each case was described in terms of either presence or absence of a symptom, and either presence or absence of a disease. The nurses' task was to decide whether there was a contingency between the symptom and the disease. The cases were prepared such that, irrespective of the disease, the symptom was present for roughly 70% of the time. Despite the objective lack of contingency between the symptom and the disease, an overwhelming majority of the nurses claimed that there was a relationship between them. The single best predictor for 'finding the relationship' was the number of cases where both the symptom and the disease were present. The nurses focused primarily on the rote frequency of 'present-present' cases, and claimed a relationship when they observed a sufficient number of such cases. Technically, they need to take the relative proportion of all possibilities into account to draw the most reasonable conclusion. This bias has been widely replicated (e.g., Shaklee and Tucker, 1980). It may be claimed that intuitive responses to percentage information are sometimes fallible, especially when such information is presented as relative frequencies.

What cognitive mechanism could underlie these observations? These results show simultaneous sensitivity to the rote frequency and insensitivity to the total number, regarding how the relative frequency is assessed. The former and the latter judgmental tendencies could be interpreted as manifestations of 'anchoring and adjustment' and 'base-rate neglect', respectively.

Anchoring and adjustment (Tversky and Kahneman, 1974) is a mental process, whereby a judge picks up a numerical clue from her/his judgmental task. S/he uses the clue as an anchor in making their response. Such processes may lead to systematic biases because adjustment is typically insufficient. It has been shown that anchoring processes mediate various numerical judgments, such as support for group decision outcomes (Allison and Beggan, 1994), preferential choice (Goldstein and Einhorn, 1987), and the number of dots in random patterns (Sawyer and Wesensten, 1994). In the examples of McFarland and Miller, as well as Denes-Raj and Epstein, the judgmental tendencies could be interpreted as showing that rote frequency served as an anchor, thereby influencing subsequent judgment.

In turn, base-rate neglect is a judgmental tendency that people underutilize relevant information about population statistics and instead overutilize other salient information (Kahneman and Tversky, 1973). For instance an intelligent student who lacks creativity, has little feel or sympathy for others but has a strong moral sense, is judged as more probably an engineering major than a humanity major. In making these judgments, people tend to overlook the fact that humanity majors are more common in the student population. Formally, the probability judgment should at least partly reflect the humanity–engineer ratio in the population, but typically that ratio is ignored. Although base-rate neglect is a robust phenomenon, it ceases to occur when the task emphasizes the relevance (such as causality) of the base-rate in producing the event under assessment (Ajzen, 1977; Hewstone, Benn and Wilson, 1988). Yet, in the risk assessment of 'cancer killing 24.14 out of 100', it is cancer, not the total being 100, that kills 24.14. Hence, it is expected that people are relatively insensitive to the base-rate in such risk assessments.

Now the prediction for the current study can be drawn. The anchoring heuristic and base-rate neglect jointly expect that a risk would be judged as more serious when the deaths were expressed by larger rote frequencies (e.g., 1,286 out of 10,000) than smaller rote frequencies (24.14 out of 100). This prediction may seem counterintuitive, suggesting that a risk with a 12.86% mortality may be judged as more dangerous than 24.14%. Concretely, I presented frequency estimates of risks in different frequencies and percentages that had been reported in previous research (Yamagishi, 1994a). The prediction was that the degree of perceived riskiness would vary according to the number of deaths presented, irrespective of the total possible deaths (such as 100 or 10,000). Moreover, I examined whether individual participants would exhibit consistent risk ratings across different methods of making judgments. Previous findings showed that, even though frequency judgments were affected by the particular method of judgment elicitation, the rank-ordering relationship among risks was well preserved across different methods (Yamagishi, 1994b). Thus, it was noted that 'probably it is in this relative sense that risk perception can be meaningfully measured' (p. 663). The following experiments investigated whether such a claim would remain adequate when participants rated the degree of riskiness instead of the frequency of deaths.

EXPERIMENT 1

Method

Participants

Fifty-two undergraduates at the University of Washington participated to earn extra course credits.

Material

Based on Yamagishi (1994a), 11 well-known causes of death were chosen. For each cause, the following list shows two percentages. The numbers denote the mean estimated percentages in Yamagishi's narrow and wide conditions, respectively. The events were: asthma (4.59,0.87); bronchitis (5.24, 1.07); cancer (24.14, 12.86); heart diseases (23.94, 15.12); HIV infection (12.55, 7.35); homicide (13.73, 4.87); influenza (5.85, 1.41); motor vehicle accidents (17.98, 8.93); pneumonia (7.55, 1.96); suicide (9.30, 3.76); and tuberculosis (5.90, 1.57). These events were presented randomly, along with other filler tasks, in questionnaire booklets. Each booklet used one of the four methods of presentation described below.

Design

A within-subject design was used. The independent variable was the format of the mortality presentation, which had four levels. The conditions differed in how the judged frequencies in Yamagishi's narrow and wide manipulations were presented. The conditions are illustrated using judged mortality rates of cancer as an example. In the LW (Larger frequency within Wide range) condition, the cause was presented with the judged frequency for Yamagishi's narrow condition with respect to 10,000 people (e.g., cancer kills 2,414 people out of 10,000). In the SW (Smaller frequency within Wide range) condition, the judged frequency for the wide manipulation was used (e.g., 1,286 people out of 10,000). In the LN (Larger frequency within Narrow range) condition, the judged percentage for the narrow manipulation concerning 100 people was used (e.g., 24.14 out of 100). Finally, in the SN (Smaller frequency within Narrow range) condition, the judged percentage for the wide manipulation was used.

The dependent variable was ratings of riskiness on a 26-point scale, ranging from 0 (*no risk at all*) to 25 (*maximal possible risk*).

Procedure

Data were gathered in a group setting. The experiment consisted of four sessions. There was a 7-day interval between the consecutive sessions. In any one session, each participant was randomly assigned to only one of the four conditions.

Before the 11 death causes were presented in the LN and SN conditions, subjects read the following: 'Shown below is a list of causes of death. For each cause, the number of people who die of the particular cause is estimated. The estimation is the number of deaths per 100 people in the public every year. For each cause, please rate how risky it appears to you'. In the LW and SW conditions, '100' was substituted with '10,000'. After this instruction, the death causes were listed. Participants were instructed to imagine playing a Russian roulette with no bullets loaded when s/he rated 'no risk at all', and playing a Russian roulette with fully loaded bullets when s/he rated 'Maximal possible risk'.

To ensure that participants' response tendencies did not vary across the four sessions (i.e., to check whether sequence effects occurred across four sessions), two questions were asked prior to rating the risks listed above. In every session, participants were told that 'statistics show that, in the US, roughly 100 people die from tornadoes each year'. Then participants rated tornadoes using the response scale described above. Next, they were instructed that 'statistics show that, in the US, roughly 9 people die of smallpox vaccination each year'. Likewise, participants rated the riskiness of smallpox vaccination. Notice that these questions did not vary the possible range of deaths (i.e., 100 or 10,000).[1] Regarding only tornado attacks and smallpox vaccinations, the participants always responded to the same question across the four sessions.

The order of administering each condition was randomized across the participants. Within each session, the order of presenting the risks was randomized across the participants.

Prediction

It was predicted that the judged riskiness would follow the rank-ordering relationship of the number of deaths presented to participants in each condition. Let est(·) denote the risk rating in each condition. The prediction is equivalent to anticipating that $\text{est(LW)} > \text{est(SW)} > \text{est(LN)} > \text{est(SN)}$. This relationship (${}_4C_2 = 6$ inequalities)[2] was expected for each cause. Overall, there were $6^* \ 11 = 66$ predicted inequalities among the total rank-ordering of judged riskiness.

Results and discussion

Two-tailed statistical tests were used in all data analyses in this paper.

---

[1] A reviewer pointed out that this procedure may have introduced a priming effect, thereby drawing the participants' attention heavily onto the absolute numbers. However, in pilot studies, data were collected without asking about smallpox vaccinations and tornado attacks (the rest of the procedure was identical to Experiment 1). Results from the pilot data showed essentially the same trend as those shown in Table 1. Therefore it seems unlikely that such priming effects were present.

[2] Other standard algebraic notations include C_2^4 and $\binom{4}{2}$.

Group data analysis

The mean ratings of the riskiness of tornado attacks from the four experimental sessions showed no significant differences (grand mean = 5.56, *F*(3, 153) = 0.25, *MSe* = 5.74). Likewise, the mean ratings for smallpox vaccination showed no differences (grand mean = 3.24, *F*(3, 153) = 1.29, *MSe* = 2.94). Therefore, I assume that the data were not contaminated by sequence effects across the experimental sessions.

Table 1 shows the mean risk rating for each condition for the 11 causes of death. For each cause, the numbers of deaths provided in stimuli were the greatest in the LW condition, followed by SW, LN, and SN. For most causes, the pattern in mean judged riskiness follows the prediction, namely est(LW) > est(SW) > est(LN) > est(SN). The exceptions were found with asthma and tuberculosis, where the data showed est(LN) < est(SN). These 2 were the only violations of the 66 predicted inequalities among the mean judged risks.

For each cause, an omnibus *F*-test was performed across the four conditions. In addition, a planned contrast was calculated between the SW and LN conditions. The contrast tested whether riskiness was more predictable from rote frequencies than percentages (e.g., was cancer rated as riskier when it was described as 'kills 1,286 out of 10,000' than as 'kills 24.14 out of 100?'). The '*F*' column in Table 1 shows that all the *F*-tests showed highly significant differences among the four conditions. The planned contrasts (see the '*t*' column) showed all positive contrast values, indicating that the risk rating was higher in the SW than in the LN condition. Of the 11 *t*-tests, 7 reached statistically reliable differences. The aforementioned violations of the predicted rank-orderings failed to show statistical significance by post-hoc tests.

Analysis of individual participant's responses

Spearman's rank correlation coefficients of risk ratings among the four conditions were calculated for each participant. Figure 1 shows notched boxplots[3] (McGill,

Table 1. Mean risk rating on a 26-point scale and test statistics in Experiment 1

	LW	SW	LN	SN	*F*	*MSe*	*t*
Asthma	5.62	3.95	3.00	3.13	12.37***	6.08	2.94**
Bronchitis	6.52	4.90	3.63	3.60	14.04***	7.01	3.39**
Cancer	12.21	10.79	8.69	8.19	14.66***	12.40	4.07***
Heart diseases	11.75	11.35	8.33	8.31	13.90***	13.12	5.37***
HIV	10.02	9.35	7.33	6.73	13.12***	9.84	4.37***
Homicide	10.23	7.25	6.69	6.13	18.80***	9.23	1.15
Influenza	6.02	4.77	3.44	3.33	10.38***	8.06	2.81**
MVA	10.81	8.37	7.88	7.81	8.69***	12.00	0.97
Pneumonia	6.44	4.33	3.90	3.85	14.17***	5.53	1.59
Suicide	7.15	5.65	4.85	4.61	10.19***	6.75	2.22*
Tuberculosis	5.60	4.12	3.58	3.77	8.07***	5.40	1.82

*, **, and *** denote the significance levels of 0.05, 0.01, and 0.001, respectively.
Motor vehicle accident is abbreviated as 'MVA'.

---

[3] In notched boxplots, the top and bottom horizontal lines of the box show the interquartile range, and the middle horizontal line the median. The oblique lines (notch) extend to the 95% confidence limits for the median.

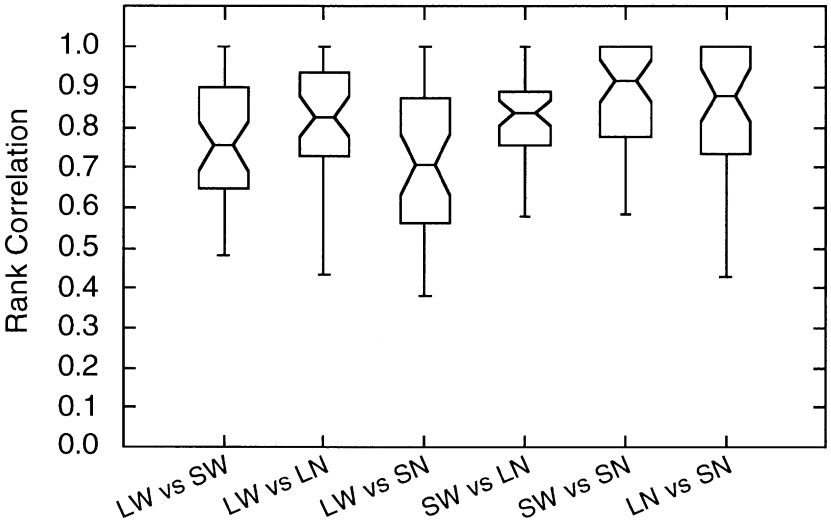

Figure 1. Notched boxplots of correlation coefficients among risk ratings (Experiment 1).

Tukey and Larsen, 1978) of the correlation coefficients. The correlation coefficients scattered mostly above 0.7, suggesting the order-preservative nature of the risk ratings across the LW, SW, LN, and SN conditions.

Thus, data were mostly consistent with the prediction that anchoring and base-rate neglect would mediate the rating of riskiness. Also, Figure 1 indicated that the relationship that one risk was judged as more or less dangerous than another was well maintained across the four experimental conditions. These results maintain coherence to earlier findings (Yamagishi, 1994a,b) in the following sense: first, numerical risk judgments are prone to applicable biases. Still, risk perception may be properly measured as *relative information*, because the rank-ordering relationship among risks did not fundamentally vary across different rating methods.

EXPERIMENT 2

Experiment 2 attempted to replicate Experiment 1's counterintuitive result, namely est(SW) > est(LN). These conditions were compared in Experiment 2. Regarding cancer, the risk rating was expected to be higher when the mortality was presented as '1,286 out of 10,000' than as '24.14 out of 100', and similarly for the rest of the risks. In addition, individual participants were expected to produce high correlations in their risk ratings in the SW and LN conditions.

Method

Participants

Forty-one University of Washington undergraduates participated to earn extra course credits.

Material

The material was the same as in Experiment 1, except for the following differences. First, different filler tasks were used. Second, the questionnaire booklets were prepared only for the SW and LN conditions.

Design and procedure

The design and procedure were also the same as in Experiment 1. The difference was that the independent variable had only two levels (SW and LN), and therefore the experiment met for two sessions.

Results and discussion

Group data analysis

The mean ratings of the riskiness of tornado attacks from the two experimental sessions showed no significant differences (grand mean = 3.85, $t(40) = 1.19$, $SEM = 0.616$, ns). Likewise, the mean ratings for smallpox vaccination showed no differences (grand mean = 3.71, $t(40) = 0.31$, $SEM = 0.469$, ns). Therefore, I assume that the participants' response tendencies did not vary across the experimental sessions.

Table 2 shows the mean risk ratings for the SW and LN conditions. The mean rating shows that est(SW) > est(LN) for every cause of death. A paired t-test was performed for each risk. The 't' column shows that the differences were significant toward the predicted direction for the nine causes (the exceptions were asthma and suicide).

Supplementary analysis

Although the results from the tornado and smallpox questions suggest little sequence effect, there remain possibilities for carry-over effects across the sessions. Thus, the mean ratings were compared between the 21 participants assigned to the SW condition in the first session and the remaining 20 participants assigned to the LN condition in the first session. This between-subject analysis rules out possibilities for

Table 2. Mean risk rating on a 26-point scale, standard error, and t-values in Experiment 2

	SW	LN	*SEM*	t
Asthma	5.14	3.78	0.683	1.99
Bronchitis	6.88	4.22	0.748	3.55**
Cancer	12.83	7.93	0.769	6.38***
Heart diseases	13.51	7.68	0.758	7.69***
HIV	11.39	6.37	0.816	6.15***
Homicide	9.88	6.54	0.644	5.19***
Influenza	5.32	3.44	0.446	4.21***
MVA	11.61	7.34	0.844	5.06***
Pneumonia	5.83	4.39	0.440	3.27**
Suicide	6.23	5.15	0.541	2.00
Tuberculosis	4.51	3.44	0.305	3.52**

*, **, and *** denote the significance levels of 0.05, 0.01, and 0.001, respectively.
Motor vehicle accident is abbreviated as 'MVA'.

finding the est(SW) > est(LN) due to carry-over effects, if any. An analogous analysis was performed using the data from the first session of Experiment 1. Because the between-subject analysis from Experiment 1 allowed for only 13 data points per condition, details are not reported here, except to mention that the 2 between-subject analyses showed the same trend.

Table 3 shows the mean rating, standard error, and the observed *t*-values of the between-subject tests (each *t*-statistic has 39 degrees of freedom). Notice the resemblance between Tables 2 and 3. As in Table 2, every death cause in Table 3 shows that est(SW) > est(LN), although fewer causes reach statistical significance. Importantly, no risk showed the mean ratings in the opposite pattern from the prediction. Thus, the finding that est(SW) < est(LN) is not solely accountable for by carry-over effects.

Analysis of individual participant's responses

For each participant, a paired *t*-test was performed on her/his risk ratings from the SW and the LN conditions. In each test, whenever a participant's responses showed that est(SW) > est(LN), the *t*-value was positive and vice versa. The left panel of Figure 2 shows the notched boxplot of the observed *t*-values for all the participants. The dotted lines show the critical values for a two-tailed test with 10 *df*, $\alpha = 0.05$. The panel shows that est(SW) > est(LN) was observed with the majority of the participants. About 75% of the participants showed the est(SW) > est(LN) pattern with statistical significance. Moreover, there was no case where the risk rating for the SW condition was significantly less than the LN condition. Consistent with the group data analysis, a great majority of the individuals showed that est(SW) > est(LN).

For each participant, Spearman's rank correlation coefficient was calculated between his/her risk ratings in the SW and the LN conditions. The right panel of Figure 2 shows a notched boxplot of the correlation coefficients. Like the results from Experiment 1, most of the correlation coefficients are distributed above 0.85. Thus, the rank ordering of riskiness among the 11 causes of death maintained consistency between the SW and the LN conditions.

Table 3. Mean risk rating, standard error, and *t*-values in the between-subject analysis, Experiment 2

	SW	LN	*SEM*	*t*
Asthma	5.38	3.55	1.289	1.42
Bronchitis	6.57	3.95	1.420	1.85
Cancer	12.19	7.60	1.767	2.60*
Heart diseases	13.62	7.30	1.795	3.52***
HIV	11.00	5.75	1.931	2.72**
Homicide	9.90	5.70	1.705	2.47*
Influenza	4.71	3.50	1.265	0.96
MVA	11.86	6.60	1.562	3.36**
Pneumonia	5.05	4.45	1.105	0.54
Suicide	6.62	4.30	1.561	1.49
Tuberculosis	4.19	3.35	0.965	0.87

*, **, and *** denote the significance levels of 0.05, 0.01, and 0.001, respectively.
Motor vehicle accident is abbreviated as 'MVA'.

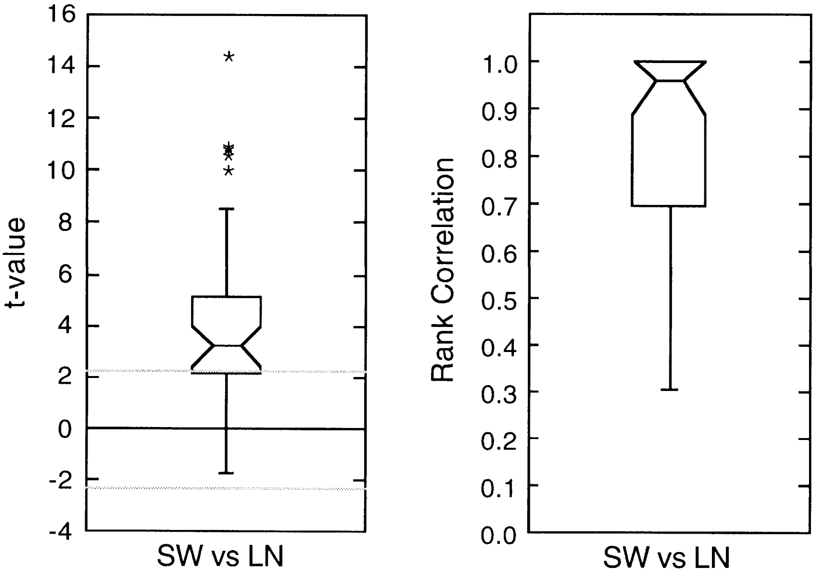

Figure 2. Notched boxplot of the observed *t*-values (left) and correlation coefficients (right) in Experiment 2.

GENERAL DISCUSSION

The results followed the prediction that anchoring and base-rate neglect would affect magnitude judgment of riskiness. The planned contrasts (Experiment 1) mostly supported the hypothesis that perceived riskiness would increase when risks were presented as relative frequencies using larger numbers (e.g., 1,286 out of 10,000) rather than percentages (e.g., 24.14 out of 100). Experiment 2 systematically replicated the response tendency. In this sense, a risk with a 12.86% mortality may be judged as more dangerous than 24.14%. Yet, individuals produced consistent rank-ordering among the risks across different methods of judgment elicitation.

These results add further to the findings by Smedslund (1963), Denes-Raj and Epstein (1994), and McFarland and Miller (1994), who demonstrated that subjective judgments were influenced when relative frequencies were presented in larger numbers. Also, the risk ratings examined here are similar to the frequency judgment task in Yamagishi (1994a,b) in that different judgment elicitation methods systematically biased the judged riskiness, yet the rank-ordering among risks was preserved across different methods. Numerical judgments on a particular risk were meaningful not at face value, but as *ranking information* relative to other risks judged in the same way.

It may appear that the use of the 'numerosity heuristic' (Pelham, Sumarta and Myaskovsky, 1995) produced the current results. When the numerosity heuristic is used, people assess quantity or probability by using the numerosity of the stimulus object as a judgmental clue. In their size-estimation task, a circle was judged as bigger when it was displayed as an array of numerous pieces in a pizza-slice shape. In their minefield problem, participants chose between two routes that differed in the numerosity of mines. Field A contained five mines and each had a 0.2 chance of explosion: field B contained ten mines, each with a 0.1 chance. Although field B provided a better probability of escaping explosion ($0.9^{10} > 0.8^5$), participants avoided field B when their cognitive capabilities were taxed. The current results may seem consistent with what the numerosity heuristic would predict in the risk assessment of relative frequencies. However, the numerosity heuristic was in effect when participants were cognitively loaded by task difficulty or other factors. The participants in the current study were not tested under such pressure. Rather, the experimental setup for the participants in the current study is similar to Pelham *et al.*'s control group, which was not put under such pressure and exhibited limited reliance on the numerosity heuristic. Hence, the numerosity heuristic has difficulties as an alternative explanation for the current results.

An analogy should be noted between the current study and a well-known anomaly in preferential choice, namely 'preference reversals' (Slovic and Lichtenstein, 1983). Preference reversal refers to changes in preferential ordering among choice options, depending upon how preferences are expressed. Particularly, provided with a choice between two gambles with comparable expected values, a decision maker tends to prefer a higher-probability gamble in choice, while s/he tends to rate the other gamble more favourably in bidding (Lichtenstein and Slovic, 1971; Lindman, 1971). Although current literature shows that the preference reversals may occur through a variety of cognitive processes, anchoring and adjustment has been one of the standard explanations (Ganzach, 1996): preference reversal reflects overpricing of the higher-payoff gamble, and such pricing occurs because decision makers anchor to the high amount of payoff and then adjust insufficiently to reach a bidding price. Superficially, the current experiments and preference reversal research are analogous because both document discrepancies between magnitude judgments on logically equivalent objects. Deeply, the two lines of research exemplify how anchoring and adjustment underlie a variety of judgmental biases. Further research is needed to determine the biases that are mediated by anchoring-adjustment processes.

It should be noted that the current results were contrary to an earlier speculation (Yamagishi, 1994a): it had been overlooked that the same percentage could be represented as rote frequencies in a variety of different ways. Consequently, it was suggested that perceived riskiness would be affected mainly by percentage. The current results showed the opposite pattern, and hence the early speculation was not supported.

Finally, concerning the asbestos example, the following may be an implication: presenting the risk to the public by teaching the 1/100,000 premature death rate may be misleading, because such information may trigger different impacts depending on equivalent but superficially different methods of presentation (e.g., 100/10,000,000 or 0.001/100). Yet, the risk could be presented as a comparison in probabilities of premature deaths due to being a pedestrian hit by a car, tobacco smoking, diagnostic X-rays, bicycling, lightning, and asbestos in school buildings. An effective risk communication may be achieved by saying that the probabilities of premature deaths due to these risks follow a descendent order. That way, lay people and policy makers might ask themselves whether they should pour millions of dollars into a risk, such as asbestos, that is less dangerous than being a pedestrian or bicycling.

REFERENCES

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, **35**, 303–314.

Allison, S. T. and Beggan, J. K. (1994). Estimating popular support for group decision outcomes: an anchoring and adjustment model. *Journal of Social Behavior and Personality*, **9**, 617–638.

Anonymous (1995). The asbestos panic attack. *U.S. News & World Report* (Feb. 20), 61–63.

Denes-Raj, V. and Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of Personality and Social Psychology*, **66**, 819–829.

Fischhoff, B. and MacGregor, D. (1983). Judged lethality: how much people seem to know depends upon how they are asked. *Risk Analysis*, **3**, 229–235.

Fischhoff, B., Slovic P. and Lichtenstein, S. (1981). Lay foibles and expert fables in judgments about risks. In T. O'Riordan and R. K. Turner (Eds.), *Progress in resource management and environmental planning* (pp. 161–202). New York: Wiley.

Ganzach, Y. (1996). Preference reversals in equal-probability gambles: a case for anchoring and adjustment. *Journal of Behavioral Decision Making*, **9**, 95–109.

Goldstein, W. M. and Einhorn, H. J. (1987). Expression Theory and the preference reversal phenomena. *Psychological Review*, **94**, 236–254.

Hewstone, M., Benn, W. and Wilson, A. (1988). Bias in the use of base rates: racial prejudice in decision-making. *European Journal of Social Psychology*, **18**, 161–176.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.

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

Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology*, **89**, 390–397.

McFarland, C. and Miller, D. T. (1994). The framing of relative performance feedback: seeing the class as half empty or half full. *Journal of Personality and Social Psychology*, **66**, 1061–1073.

McGill, R., Tukey, J. W. and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, **32**, 12–16.

Pelham, B. W., Sumarta, T. T. and Myaskovsky, L. (1995). The easy path from many to much: the numerosity heuristic. *Cognitive Psychology*, **26**, 103–133.

Sawyer, T. F. and Wesensten, N. J. (1994). Anchoring effects on judgment, estimation, and discrimination of numerosity. *Perceptual and Motor Skills*, **78**, 91–98.

Shaklee, H. and Tucker, D. (1980). A rule analysis of judgments of covariation between events. *Memory and Cognition*, **8**, 459–467.

Slovic, P. and Lichtenstein, S. (1983). Preference reversals: a broader perspective. *American Economic Review*, **73**, 596–605.

Slovic, P., Layman, M., Kraus, N., Flynn, J., Chalmers, J. and Gesell, G. (1991). Perceived risk, stigma, and potential economic impacts of a high-level nuclear waste repository in Nevada. *Risk Analysis*, **11**, 683–696.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, **4**, 165–173.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, **185**, 453–458.

Yamagishi, K. (1994a). Consistencies and biases in risk perception. I. Anchoring process and response-range effect. *Perceptual and Motor Skills*, **79**, 651–656.

Yamagishi, K. (1994b). Consistencies and biases in risk perception. II. Individual subjects' performance. *Perceptual and Motor Skills*, **79**, 659–663.