Journal of Risk and Uncertainty, 14:301–309 (1997)
© 1997 Kluwer Academic Publishers

Confusion of Relative and Absolute Risk in Valuation

JONATHAN BARON
Department of Psychology, University of Pennsylvania, 3815 Walnut St., Philadelphia, PA 19104-6196.
(E-mail) baron@psych.upenn.edu.

Abstract

Subjects were less willing to pay for government medical insurance for diseases when the number of people who could not be cured was higher, holding constant the number who could be cured. In a second experiment, willingness to pay (from a hypothetical government windfall) for risk reduction was unaffected by whether the risk was described in terms of percentage or number of lives saved, even though subjects knew that the risks in question differed in prevalence. These results are consistent with the findings of Fetherstonhaugh et al., Jenni and Loewenstein, and others. I suggest that these results can be explained in terms of a general tendency to confuse proportions and differences, a confusion that is analogous to other confusions of quantitative dimensions in children, adults, the news media, and perhaps even the epidemiological literature.

Key words: relative risk, value of life, contingent valuation

JEL Classification: J17

Introduction

People are confused about quantities. The problem begins in childhood. Small children confuse length and number, so that, when we ask them to compare two arrays for length or number, they will answer with little regard to which question we asked (Baron, Lawson, & Siegel, 1975). In general, they tend to answer as if they were asked about length when the number of items in the array is larger than they can count easily, so they are correct when asked which array is longer but incorrect if asked which array has more items. When the number of items is small, e.g., 3–5 items, they are correct when asked which has more but incorrect when asked which is longer. Training to answer answer length question correctly leads to increases in accuracy at length questions and decreases in accuracy at number questions, and training to answer number questions correctly (with more numerous arrays) has the opposite effects.

Similar quantitative confusions are found in older children and adults. Older children are confused about probability, so that they answer probability questions according to frequency rather than relative frequency (Piaget & Inhelder, 1975). Another example of this kind of confusion is found in probability judgment in adults (Kahneman & Tversky, 1972). When subjects are given a sample drawn from one of two distributions of red and black chips, with equal prior probability, the posterior probability that they assign depends on the proportion of red chips in the sample. According to Bayes's theorem, the posterior probability should depend on the difference between the number of red and black chips. Although the exact dependence on the difference is not intuitively obvious, the point of interest here is that subjects' judgments are independent of the sample size. They use the ratio inappropriately.

Even when people, in some sense, know which quantitative dimension is relevant, they are influence by other dimensions. The tendency to use frequency when probability (a proportion) is relevant continues into adulthood, so that many people prefer to bet on an urn with 9 winning chips out of 100 rather than an urn with 1 out of 10 (Denes-Raj & Epstein, 1994). In a related demonstration, subjects were more suspicious that a male supervisor was biased against women when the highest score on a test he graded was obtained by the single male out of 10 employees than when the highest score was obtained by one of 10 males out of 100 employees (Miller et al., 1990). Thus, adults continue to be influenced by frequency even when they know that relative frequency is relevant.

Similar confusions occur in news media and perhaps even in scientific journals. Newspapers often tell us that "inflation increased by 2.9%" when they mean that prices increased by this much. The literature on risk effects of pollutants and pharmaceuticals commonly reports relative risk, the ratio of the risk with the agent to the risk without it, rather than the difference. Yet, the difference between the two risks, not their ratio, is most relevant for decision making: if the risk is minuscule, a high relative risk still means very little.

When people attend to the ratio of small risks, they can exaggerate the importance of changes when risks are small. Estrogen replacement therapy, for example, increases the risk of endometrial cancer several fold, but the risk is tiny. Despite the small absolute effect, early opposition to estrogen replacement may have focused on it (Hershey & Baron, 1987; now the effect is reduced by addition of progesterone, but this, too, may be an overreaction). Stone, Yates, and Parker (1994) found that presentation of relative risk information, as opposed to full information about the two absolute risks involved, made people more willing to pay for safety when risks were small.

Another kind of quantity confusion occurs in the evaluation of goods. When asked how much they are willing to pay for risk reductions or other goods, people are often willing to pay the same amount of money for different quantities of what they are evaluating (e.g., Jones-Lee, Loomes, & Phillips, 1995). Baron and Greene (1996) replicated this result but also found that subjects gave the same willingness to pay (WTP) for different quantities when they were asked their willingness to pay per unit of the good. It is almost as if their original answers were per unit rather than for the total amount, even though they were asked about the total. McDaniels (1988) reported similar results for valuation of risk reductions.

Analogous results appear in ratings of benefit from public expenditures. Kemp and Willetts (1995) asked subjects to rate the value of government services in New Zealand, including items varying in the money spent on them, e.g., government retirement income (NZ$4,314 million), universities (NZ$577 million), and the New Zealand Symphony Orchestra (NZ$8 million). Subjects rated "how useful or worthwhile" each item was to New Zealand, or "how much value it would be to New Zealand." One group of subjects rated the total utility of each service and the utility of a 5% increase in spending (marginal utility of percent). The other group rated the utility per dollar spent and the utility of each extra dollar spent on the program (marginal utility of dollars). Subjects were not told the present cost of each service, but it would be reasonable for them to think that the amounts differed substantially.

Correlations of the four different kinds of ratings (based on means across subjects) were high, and unaffected by cost. We might expect that total utility would correlate highly with cost, while per-dollar utility would not correlate much with cost. (Ideally, according to economic theory, marginal utility per dollar should be equal for all services.) In fact, both total and per-dollar ratings correlated moderately across services with log(cost), $r = .55$ in both cases. Subjects failed completely to distinguish between total utility and utility per dollar.

Marginal-utility-of-dollars ratings should be uncorrelated with cost or with any other ratings, although we might expect positive correlations between cost and marginal-utility-of-percent ratings. The correlation of marginal-utility-of-dollars with cost was .38, and the correlation of marginal-utility-of-percent was .41. Marginal ratings also correlated over .90 with total ratings in both conditions, and they correlated .99 with each other (across programs). In sum, subjects did not distinguish among different kinds of questions. Their ratings make more sense if we think of them as ratings of total utility (or perhaps of utility of the first dollar spent), since the ratings were correlated with cost. These findings thus seem to be another example of contamination by an irrelevant factor, which could be cost or the total scope of the program.

In the experiments reported here, people attend to ratios as well as differences when they should attend only to differences. Similar results are found by Fetherstonhaugh et al. (1997) and by Jenni and Loewenstein (1997). The present experiments use WTP, thus extending the general conclusion to that measure. In addition, I argue that these results may be analogous to other confusions of the sort just reviewed.

Experiment 1

In evaluating the benefit of a program, people may attend not only to the actual benefits but also to the potential benefits, the total problem. People may be less willing to contribute when their contribution is a “drop in the bucket,” even when the effect of the contribution is held constant. Unger (1996) argued that such “futility thinking” underlies the common intuition that failing to help the world’s hungry people is not a great moral offense: although we could save many lives at little cost per life, each individual (or even each nation) cannot make more than a dent in the whole problem.

This experiment asks subjects about their WTP for coverage for lifesaving medical treatment as a function of both the number of lives saved and the total number of people with the disease in question. At issue is whether people are influenced by the ratio of lives saved to the total as well as by the number of lives saved.

Method

The items used were the first eight from a longer questionnaire, the results of which were fully consistent with those reported.

Subjects were instructed: "Suppose that, in a few years, the U.S. government decides that everyone will have a standard medical insurance package. Some costs would not be covered. Extra insurance for these costs will be expensive, and most people will settle for the standard package. The average cost of the package would be about $1,000 per year for each person. Payment will be distributed so that the rich pay more than $1,000 and the poor pay less.

"Once basic coverage is decided on, a few procedures are considered borderline. We ask you how much extra you would be willing to pay for each of the procedures listed below. The extra cost of each procedure will be added to the $1,000 average cost of the standard policy."

In addition, subjects read that the decision about whether a treatment would be covered would be based on a comparison of the median WTP to the cost of the treatment, and: "Each numbered item concerns a different disease. All figures refer to the U.S. only. When a treatment benefits only some of the people with a disease, that is because the disease has different forms."

The first item for half the subjects read: "1,000 people die from this disease each year. Their average age is 60. How much are you willing to pay to cover a treatment that will save the lives of 900 of these people?" The next items asked about curing 90 out of 100, 900 out of 10,000, and 90 out of 1,000, and then these four items were repeated. For items 1, 2, 7, and 8, the average age was 60; for the rest, it was 40. The age manipulation was included to give subjects license to decide items independently. But the items to be compared involved low vs. high cure rates (90 vs. 900) and low vs. high ratios (.09 vs. .90). For the other half of the subjects, order was reversed; order did not affect the results.

Subjects were 95 students from the University of Pennsylvania and Philadelphia College of Pharmacy and Science, paid by the hour for completing this questionnaire and others. An additional nine subjects failed to complete the questionnaire, and an additional four subjects were not counted because they gave WTPs of 0 for at least half of the eight items (making calculation of logs suspect).

Results

Although WTP was affected by lives saved (90 vs. 900), it was also affected by the ratio of those saved to those with the disease (.09 vs. .90). Age had an effect as well. The geometric means were, respectively: $143 for 900 saved and $74 for 90 saved ($t = 9.03$, $p < .0005$ based on log means); $129 for a .90 ratio and $89 for a .09 ratio ($t = 5.29$, $p < .0005$); and $132 for age 40 and $88 for age 60 ($t = 6.24$, $p < .0005$).

WTP was somewhat unresponsive to quantity (900 vs. 90 lives saved), as others have found: the WTP ratio was less than 2, much smaller than the tenfold difference in lives saved. The effect of ratio (.09 vs. .90) was smaller than that of lives saved, but still substantial. These two effects (measured as differences of logs) were correlated negatively across subjects ($r = -.22$, $p = .031$), so it would appear that subjects differ in whether they attend to the proportion of lives saved or to the number.

In sum, subjects attend to proportion of lives saved as well as number in deciding their WTP.

Experiment 2

The second experiment, based on the idea of the Kemp and Willetts (1995) study described earlier, asked subjects about their willingness to pay for government programs to save lives from several major causes of death (culled from Gardner & Hudson, 1996). Subjects indicated WTP for a 5% reduction in that cause of death (WTP-percent) and a reduction of 2,600 Americans per year (WTP-lives). They also gave their own estimates of the number of Americans who died from each cause (prevalence) and they rated the relative importance of preventing death from each cause.

These data allow us to determine whether each subject distinguishes between percentage changes and absolute changes, and, if not, whether WTP is based implicitly on proportional or absolute change. If subjects think in terms of absolute change, then their prevalence judgments should correlate with WTP-percent and not with WTP-lives, especially when importance is partialled out.

Method

Twenty-nine subjects, solicited as in Experiment 1, answered a questionnaire, which began: “Here is a list of causes of death from the Centers for Disease Control. Suppose that the U.S. government had a budget surplus of $100,000,000,000 and that this money was to be allocated to programs to save lives, of all sorts. Some of the money could be allocated to reduction of death rates from the following causes. Half of the subject were asked first to go through the list and indicate, “How much money would you allocate to reducing each of these causes by 5%, if you were sure that the program would be effective?” Then they indicated, in a new list, “How much money would you allocate to reducing each of these causes by 2,600 Americans per year, if you were sure that the program would be effective? (The population of the U.S. is about 260,000,000, so this is about 1 out of every 100,000 Americans.)” The other half did these two tasks in reverse.

All subjects then indicated, in another list, their estimate of the annual death rate (prevalence) from each cause per 100,000 (with the total rate given in the first row), and, in a final list, they answered the question, “Some deaths may be more important to prevent than others. Please rate the relative importance of preventing each kind of death. Give 100 to the most important and rate the others in proportion to their importance. (Ties are allowed.)” (Two subjects ranked the causes instead, so their responses were reversed for computation of correlations.)

In sum, each subject indicated, for each risk, WTP according to percent (WTP-percent) and according to lives (WTP-lives), death rate, and importance. For purposes of comparison, I also computed for each subject the responses to the WTP-lives question that could be inferred from the subject's responses to WTP-percent and prevalence, WTP-lives-inferred.

Results

Subjects generally failed to distinguish between the two expressions of quantity, WTP-lives and WTP-percent. The median correlation (computed within each subject across the 18 causes) between the two measures was .80, with a correlation of 1.00 in 26% of the subjects. These subjects apparently saw no reason to make any distinction between WTP for a percentage change and WTP for a given change in number of lives, even though they certainly understood (as indicted by their responses to the prevalence list) that rates differed for different causes. This result is a form of complete insensitivity to quantity.

Both kinds of WTP responses are higher when the subject judged the cause to be more prevalent. The mean correlations with the subject's prevalence judgments are .50 for WTP-percent and .44 for WTP-lives. Both correlations are significantly positive ($p < .0005$, tested across subjects), but they do not differ significantly across subjects. The correlation with prevalence suggests that both measures are being treated to some extent as if they were WTP for a proportional change. There would be no reason for a correlation between WTP-lives and prevalence unless importance just happened to correlate with prevalence. In fact, it does (mean $r = .48$), but, when importance is partialled, the correlations are still both significant (and not significantly different).

On the other hand, it also appears that WTP is not completely responsive to judged prevalence. If we divide WTP-percent by prevalence, we get a WTP-lives-inferred measure, representing WTP for a fixed number of lives. This measure should be uncorrelated with prevalence, if the WTP-percent measure were fully responsive to prevalence. In fact, WTP-lives-inferred was negatively correlated with judged prevalence (mean $r = -.28$, $p < .0005$ across subjects).

It thus appears that subjects were striking some compromise between taking prevalence into account and ignoring it, but equally so for both WTP measures. They were willing to pay more for more common disorders, but not in proportion to their prevalence. This result was also found by McDaniels (1988) for national expenditures and for a WTP survey.

It should be noted that three subjects stood out from the general pattern, showing high correlations between prevalence and WTP-percent and much lower correlations between prevalence and WTP-lives, as would be expected if they were responding differently to the two questions. Although most subjects did not make this distinction, it appears possible to do so.

Importance also affected WTP. WTP-percent and WTP-lives both correlated with importance (mean $r = .55$ for both; $p < .0005$). The correlations between importance and WTP remaining highly significant when judged prevalence was partialled out for each subject. Subjects thus regard some deaths as more worth preventing than others.

Table 1 shows the medians of subjects' responses, and the CDC figures for death rates. All correlations between one column and another column were significant across the items at $p < .02$ or better. The correlation of WTP-percent and WTP-lives was 0.96. The typical subject did not distinguish the two questions.

Discussion

People fail to distinguish fully among measures of effects based on differences or ratios. In particular, some subjects evaluate health benefits in terms of the proportion of lives saved rather than the number. People are excessively influenced by the prevalence of a risk when they are asked how much of the national budget should be spent to save a fixed number of lives, but they are also undersensitive to the prevalence of the risk when they are asked how much should be spent to save a percentage of lives.

The percentage question is more like asking how much of the budget should be spent on each type of risk reduction. In general, we would expect risk reduction efforts to reduce deaths by a percentage that is roughly independent of the number of expected deaths. If this is true, and if national budgets are responsive to the undersensitivity to prevalence found here, we would expect excessive allocations to small risks. Budget allocations should likewise be too sensitive to the proportion affected rather than the total number.

Table 1. The causes of death used in Experiment 2, the CDC figures (age adjusted) on their prevalence per 100,000, and the medians of subjects responses (WTP in billions).

Cause of death	CDC	prevalence	importance	WTP-percent	WTP-lives
1 Heart	145	55	90	10	9
2 Stroke	26	30	80	5	5
3 High BP	2	24	80	5	5
4 Cancer	133	50	90	10	10
5 Chronic lung	21	20	60	5	4
6 Car accidents	16	38	70	5	5
7 Other accidents	14	25	65	3	2
8 Pneumonia/flu	13	10	50	2.5	3
9 Diabetes	12	20	65	5	2.5
10 HIV infection	14	25	85	10	10
11 Suicide	11	15	50	1	1
12 Suic./firearms	7	10	50	1	1
13 Homicide	11	20	80	5	5
14 Hom./firearms	7	16	80	5	5
15 Chronic liver	8	15	40	4	2.5
16 Alzheimer's	2	10	40	3	3
17 Drugs/alcohol	11	30	70	5	5
18 Pregnancy, etc.	7	15	60	4	5

McDaniels (1988) found such results in actual government expenditures on risk reduction as well as in survey results with a small number of risks. In particular, the implied national WTP per life saved was higher for regulations that affected very few people (e.g., workplace chemicals) than for those that affected many people (e.g., auto safety).

The normatively relevant index in Experiment 1 is salient and easy to compute. It is not surprising, then, that it plays a large role in judgments. When numbers are not provided, however, people may be more prone to evaluate programs on the basis of proportional effects rather than absolute effects. Experiment 2 may have been more subject to this effect. This effect could occur even when people think about programs on the basis of their general knowledge. For example, a research program directed at a rare disease might seem almost as valuable as one directed at a common disease, because people might think of the programs in terms of their effectiveness relative to the size of the problem that each program addresses.

An alternative explanation of some of the results described here, and other results, is that people evaluate risk reductions according to a function like that of prospect theory (Kahneman & Tversky, 1979). This explanation is favored by Fetherstonhaugh et al. (1997). Although it is possibly correct in some cases, it cannot easily explain the present results, or some of those obtained by Fetherstonhaugh et al. In particular, some comparisons of interest in Experiment 1 held constant the number of lives saved and vary only the number that might be saved. The explanation of different evaluations in these cases could be psychophysical, but it is not based on prospect theory. It must involve some influence of relative risk, even though absolute risk is what matters.

The general tendency to confuse related quantitative dimensions is pervasive in human reasoning. The fact that some of the confusions occur in children but not in adults is encouraging. If the effects of development result from education, then perhaps more explicit efforts to educate citizens about important quantitative distinctions would be worthwhile. Also helpful might be greater efforts, by reporters, government officials and authors of technical papers, to present statistics in the form most relevant to making decisions

Acknowledgement

This research was supported by N.S.F. grant SBR95-20288.

References

Baron, J., & J. Greene (1996). Determinants of insensitivity to quantity in valuation of public goods: contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*.

Baron, J., G. Lawson, & L. S. Siegel (1975). Effects of training and set size on children's judgments of number and length. *Developmental Psychology*, 11, 583–588.

Denes-Raj, V., & S. Epstein (1994). Conflict between experiential and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, 66, 819–827.

Fetherstonhaugh, D., P. Slovic, S. Johnson, & J. Friedrich (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty*.

Gardner, P., & B. L. Hudson (1996). Advance report of final mortality statistics, 1993. *Monthly Vital Statistics Report*, 44(7), supplement. Centers for Disease Control and Prevention, National Center for Health Statistics.

Hershey, J. C., & J. Baron (1987). Clinical reasoning and cognitive processes. *Medical Decision Making*, 7, 203–211.

Jenni, K. E., & G. Loewenstein (1997). Explaining the "identifiable victim effect." *Journal of Risk and Uncertainty*.

Jones-Lee, M. W., G. Loomes, & P. R. Philips (1995). Valuing the prevention of non-fatal road injuries: Contingent valuation *vs.* standard gambles. *Oxford Economic Papers*, 47, 676 ff.

Kahneman, D., & A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.

Kahneman, D., & A. Tversky (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 263–291.

Kemp, S., & K. Willetts (1995). Rating the value of government-funded services: comparison of methods. *Journal of Economic Psychology*, 16, 1–21.

McDaniels, T. L. (1988). Comparing expressed and revealed preferences for risk reduction: Different hazards and question frames. *Risk Analysis*, 8, 593–604.

Miller, D. T., W. Turnbull, & C. McFarland (1990). Counterfactual thinking and social perception: Thinking about what might have been. In M. Zanna (Ed.), *Advances in Experimental Social Psychology*, 23, pp. 305–331. Orlando, FL: Academic Press.

Piaget, J., & B. Inhelder (1975). *The origin of the idea of chance in children*. New York: Norton (originally published 1951).

Stone, E. R., J. F. Yates, & A. M. Parker (1994). Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavior and Human Decision Processes*, 60, 387–408.

Unger, P. (1996). *Living high and letting die*: *Our illusion of innocence*. Oxford: Oxford University Press.