

How bad is a 10% chance of losing a toe? Judgments of probabilistic conditions by doctors and laypeople

ANDREA GURMANKIN LEVY and JONATHAN BARON
University of Pennsylvania, Philadelphia, Pennsylvania

We presented a Web questionnaire to 139 physicians and medical researchers and 109 laypeople. The subjects made judgments of badness and importance of prevention for eight medical conditions at each of seven different probability levels. By assuming that the response to each of the 56 risks was monotonically related to transformations of the probability and of the disutility of the condition, we could assess the relative effect of probability and disutility on each subject's judgments. Physicians' judgments were more sensitive than laypeople's judgments to changes in probability. Older and female laypeople were less sensitive to probability (and correspondingly, more responsive to differences in severity among medical conditions). Laypeople varied more than physicians in their responsiveness to probability. These results point to general individual differences in the effect of probability on evaluations of medical risks. They may also provide insight into causes and noncauses of physician-patient miscommunication.

Medical decisions often involve evaluations of risks, such as a 5% chance of death or a 50% chance of sexual impotence, as a result of some treatment. Health professionals frequently communicate these risks to patients in roughly this form, a probability and an outcome, so that patients can make a decision themselves, understand why their health care provider made a recommendation, or give informed consent. Arguably, physicians and patients should evaluate such risks in terms of their expected utility (EU). If they evaluate risks in this way, they can easily translate their evaluations into decisions that best achieve their goals. If other things are equal, they will choose the best option for risk reduction if they consider reduction in terms of EU (Baron, 2004). They should, roughly, multiply the probability by the disutility of the outcome in order to arrive at an overall assessment of the seriousness of the risk. Such an approach, and many similar approaches, require an understanding of communicated probabilities and their trade-offs with the disutility (i.e., severity) of possible outcomes.

People may differ in how they respond to probability, relative to the disutility of the outcome, and health professionals may differ systematically from their patients. For example, patients may think of a very low probability (e.g., 0.1%) as quite high, thus failing to distinguish high and low probabilities, whereas physicians may make greater

distinctions because they are more familiar with probabilistic reasoning (Nisbett, Fong, Lehman, & Cheng, 1987). Physician-patient differences in judgments of the probability of some outcome may account for a host of problems in the physician-patient relationship, such as miscommunication, patients' misunderstandings of their situation and options, and patient dissatisfaction with physician communications (Collins, Clark, Petersen, & Kressin, 2002; Gurmankin, Domchek, Stopfer, Fels, & Armstrong, 2005).

In addition, departures from EU may help explain some suboptimal patient decisions, such as nonadherence to medications or to recommended screening tests, on the one hand, or excessive health care utilization, on the other. For instance, if a physician tells a 35-year-old woman at average breast cancer risk that her risk of developing breast cancer in the next 5 years is 0.3%, the physician will regard this as a very low risk, whereas the patient may perceive it to be high and may, therefore, conduct daily breast self-exams, seek frequent mammograms, and undergo biopsies of any masses or lumps, despite the extremely low probability of their being malignant.

In the present study—the first of its type, to our knowledge—we presented laypeople and physicians with simple risks (e.g., “a 10% chance of amputation of a big toe”) and asked for judgments of *badness* and of importance of prevention for each risk. We assessed how these judgments depended on the outcome and the probability. We might expect badness to correspond to expected disutility, since this is roughly what badness means. Importance of prevention corresponds more obviously than badness, since it is reasonable to assume that, other things (such as cost) being equal, we would maximize utility by acting according to importance.

A.G.L. is currently at the Dana-Farber Cancer Institute and the Harvard School of Public Health. Correspondence concerning this article should be addressed to J. Baron, Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104-6196 (e-mail: baron@psych.upenn.edu).

The normative theory of EU holds that the expected disutility of a prospect is the product of the disutility of the outcome and its probability. Thus, people's perceived disutility of a risk should be the product of the disutility of the outcome and its probability. However, the response measure need not be linearly related to the perceived disutility of the risk, and the internal representation of probability (subjective probability) may also be distorted. In making judgments, people may behave as if they converted probability statements into some number that was monotonically related to the probability, converted disutility of the outcome into another number, multiplied the two numbers, and then converted the product monotonically into a response judgment. This possibility is the basic assumption of conjoint measurement (Krantz, Luce, Suppes, & Tversky, 1971).

In the present study, in which subjects evaluated simple risks, each consisting of a single outcome and a nonzero probability, multiplication of *subjective* probability and outcome utility is indistinguishable from addition of subjective probability and outcome utility. Thus, we fit a model that assumes additivity of the two representations.

Our approach is similar to that of Tversky (1967), who carried out a similar analysis of utility and probability for paired comparisons of simple gambles for cigarettes or candy, in a prison population for which these goods had some exchange values. Tversky recovered the relation between the stated probability and the subjective probability and, likewise, the relation between utility and the number of cigarettes or candies. He found distortions of probability and nonlinear utility scales for cigarettes and candy. The probability distortions he found were later incorporated into prospect theory (Kahneman & Tversky, 1979, 2000).

Although multiplication of subjective (nonzero) probability and utility is indistinguishable from addition, it is possible to distinguish addition and multiplication when zero probabilities are included or when different tasks can be assumed to rely on the same subjective representations. In such situations, Mellers and Chang (1994) and DeKay and McClelland (1996) have found that people rate the seriousness of risks more as if they were adding than multiplying. It is also possible, as Tversky (1967) did, to determine the relationship between the probability supplied by the experimenter and the subject's subjective probability. If subjects are following EU, using probabilities as stated, subjective probability should be proportional to the logarithm of stated probability.

In an effort to avoid at least some of the distortions of probability that have been found, we presented probabilities in logarithmic steps. If people were to form a weighted sum of the logarithm of the probability and (some transformation of) the utility of the outcomes, they would appear to follow EU, in terms of the ranking of their judgments.

METHOD

We used two groups of subjects: 109 laypeople and 139 physicians. The laypeople completed a questionnaire on the World-Wide

Web for \$3 after receiving an e-mail notification sent to 500 potential subjects. The study was removed from the Web when 100 responses had been received, but 16 more came in after that, from subjects who had already started it. The laypeople ranged in age from 18 to 62 years (median, 34), and only 18% were male. Seven others were removed because they were part of a group that was distinct from other subjects in being both very fast and extremely inconsistent, leaving 109 subjects.

The physicians consisted of two subgroups. The 59 members of one subgroup had just graduated from medical school.¹ Their ages ranged from 24 to 37 years (median, 26), and 47% were male. They were paid \$5 each.

The other physician subgroup consisted of 80 members of an electronic mailing list (with 993 addresses) for academic physicians at the University of Pennsylvania School of Medicine. They were given a chance to enter a drawing for a \$300 prize for completing the study. Their ages ranged from 31 to 72 years (median, 43), and 73% were male. (Ten subjects in this group did not report their age, and 6 did not report their sex.) The median number of years since completing medical school was 16. (Eight respondents left this question blank.) The mean percentage of time devoted to clinical practice was 37%, including 1 who indicated no clinical practice, 2 for whom this could not be determined, and 6 others who did not answer the question. We included all the subjects in the analysis, because we assumed that increased familiarity with probability concepts results from the experience of working in a medical setting, as well as from clinical practice.

The questionnaire consisted of 56 risks: eight medical conditions at each of seven different probabilities. All the subjects rated the badness of each risk. The laypeople and the new medical school graduates also judged the importance of prevention. We omitted this from the version given to the academic physicians, in order to reduce the time required.

The introduction to the full version read as follows, with the second version differing only slightly:

Risk severity

This is about judgments of the severity or badness of risks. Each item describes a health condition, and a probability of getting it. The probability applies to you. **Please read the following points carefully before you begin.**

- In all cases, imagine that, if you get the condition, **it begins immediately, nothing else happens, and it is not a signal that anything else is more likely to happen.** (This is rarely true in real life, but try to imagine it.) That is, you simply change from not having the condition to having it.

- One of the "conditions" is immediate death. We ask you to compare each item to a 100% chance of immediate death.

There are 56 screens, each with two questions, and a few questions at the end. The 56 screens consist of 8 different health conditions, each presented at 7 different probabilities, in a random order.

The conditions, in order of severity (in our judgment and that of the subjects) were (with abbreviations) *wart*, a harmless, painless wart on a big toe; *toe*, amputation of a big toe; *deaf1*, complete deafness in one ear (with normal hearing in the other ear); *leg1*, amputation of one leg below the knee; *leg2*, amputation of both legs above the knees; *Blind*, complete blindness in both eyes; *BBDD*, complete blindness and deafness; *death*, immediate death.

The probabilities were 0.1% (1 out of 1,000), 0.32% (32 out of 10,000), 1% (1 out of 100), 3.2% (32 out of 1,000), 10% (1 out of 10), 32% (32 out of 100), and 100% (absolutely certain). The probabilities were spaced logarithmically, and the smallest denominator was used in all cases. The denominator for powers of 10 was lower than that for other probabilities.

The 56 screens were presented in a different random order to each subject. The two questions were, for example,

How bad is it to face a 32% (32 out of 100) chance of a harmless, painless wart on a big toe (on a scale where 0 is *not bad at all*, 100 is *as bad*

as a 100% chance of immediate death, and over 100 means worse than that)? Pick the closest:

0 1 2 3 4 5 7 9 11 14 18 23 30 38 48 62 79 100 127

How important is it to take steps to take measures to prevent completely a 32% (32 out of 100) chance of a harmless, painless wart on a big toe?

- Not important at all
- Slightly important
- Somewhat important
- Very important
- Extremely important
- As important as preventing a 100% chance of death

Again, the badness judgments were on a scale that was roughly logarithmic.

The second version was essentially the same, except for elimination of the question about prevention.

RESULTS

We will report the results for the badness question for all the subjects and the results for the prevention question for the subjects who answered it (i.e., the laypeople and the new graduates). The two physician subgroups did not differ significantly on any dependent variables and were, therefore, combined into a single physician group for analyses of the badness judgments, which both physician subgroups provided. The entire physician group did not differ significantly from the laypeople in age (median, 36 years for the physicians and 34 years for the laypeople), although these groups did differ in sex (physicians were 62% male; laypeople, 18%).

The Relative Importance of Probability

Figure 1A shows the mean badness judgments as a function of probability for all the subjects (each step is one unit on the scale we used). Each line is one of the eight medical conditions. Figure 1B shows the results for the mean

prevention response. Note that the badness judgments show a fanning-out pattern not found in the prevention judgments, perhaps because of a ceiling effect, given that 15% of all the responses to the prevention question were at the maximum of 5.

The raw badness judgments do not correspond directly to EU. If they did, the judgments for probabilities of .001, .01, .1, and 1 would differ by a factor of 10. Although the mean ratio of .01 to .001 is 7.3, the other ratios are all less than 3. The prevention judgments seem even less sensitive to probability. This departure does not mean that the subjects were violating EU, for it could also be explained by use of the response scale.

We did not attempt to explain to the subjects that the scales represented expected disutility. Thus, although we should expect both of these scales to be monotonically related to EU, we should not expect them to be linearly related. Each response scale could be some transform of EU.

Thus, we transformed the variables for each subject so that the transformed responses were (as closely as possible) a linear sum of the transformed condition weights (i.e., the weight for each medical condition) and the transformed probabilities, which we call *probability weights*. The condition weights are estimated; they are assumed to be a monotonic transform of the condition disutilities. Because some subjects made occasional internally inconsistent responses (which could not be corrected by any simple rule that we could discover), we constrained the transformations to be monotonic in both probability and condition (with condition ordered by severity, as shown in the figures, which coincides with what we intended to be the severity ranking in the eight conditions). If the transformed probabilities and condition weights are logarithms of probabilities and utilities, respectively, EU theory would imply that their sum is the log of the EU.

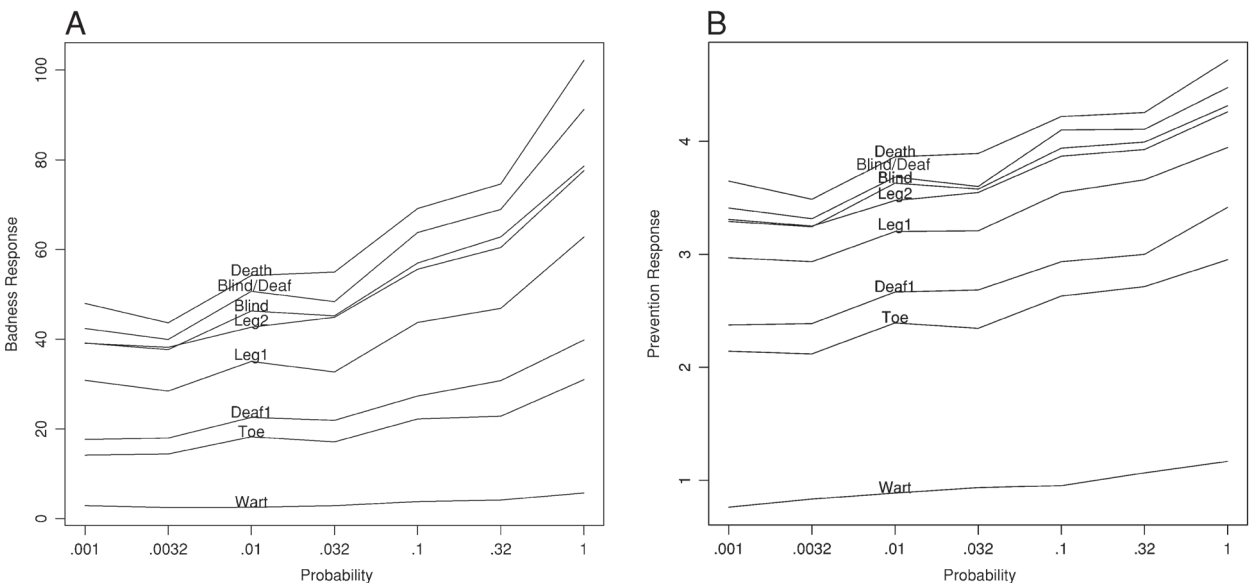


Figure 1. Responses as a function of probability and condition.

To accomplish the transformation, we used the Alternating Conditional Expectations package for R , which uses an iterative procedure to maximize the proportion of variance in the transformed ratings that is accounted for by the two other variables, transformed (R Development Core Team, 2004). The main output of this analysis, for each subject, is a list of seven weights for the probabilities, eight weights for the conditions, and numbers for all the different ratings that the subject used, so that each transformed rating was approximated by the sum of the corresponding probability weight and condition weight.

The transformations of ratings themselves were generally concave and were more concave for badness than for prevention judgments. Figures 2A and 2B show the mean transformed responses as functions of the mean (linearly) transformed probabilities, with one line for each condition. The transformation was based on the assumption of additivity, so the lines tend to be parallel.

The probability transformations are of interest in telling us how people use the probability scale (as in Tversky, 1967). In general, they were convex in terms of the logarithmic scale that we used [$t(247) = 11.41, p = .0000$, for the curvature (quadratic component), computed within each subject and tested across subjects; the degree of curvature did not depend on physician (both groups combined, vs. laypeople), age, or sex]. The result was similar for prevention [$t(167) = 4.00, p = .0001$]. We will discuss this result later.

To analyze the relative sensitivity of each subject's badness and prevention judgments to probability vs. medical condition, we took the ratio

$$\frac{\text{probability slope}}{(\text{probability slope}) + (\text{condition slope})},$$

where probability slope is the regression coefficient when transformed probability is predicted from the logarithm of

probability, and condition slope is the regression coefficient when transformed condition utility is predicted from the overall rank order of the condition badness. This ratio would be 0 if probability had no effect on judgments and 1 if condition had no effect. (We allowed negative values of the numerator, because we assumed that these resulted from random variation when subjects largely ignored probability.) This ratio is the most direct measure of the extent to which people attend to probability. It is roughly based on the rankings of condition–probability pairs.

Although badness and prevention judgments are close after this transformation, the probability sensitivities were higher for the badness judgments than for the prevention judgments [$.304$ vs. $.238$; $t(167) = 5.97, p = .0000$].²

Group and Individual Differences in Probability Sensitivity

We analyzed probability sensitivity of badness judgments as a function of the subject's being a physician, sex, age, and speed at doing the task. Speed was defined as the logarithm of the mean of the fastest half of the trials. The subjects did not know that we timed their responses. Analyses involving speed are restricted to the subjects who made both badness and prevention judgments.

In the full sample, the physicians were more sensitive to probability than were the laypeople [$.37$ vs. $.27$; $t(246) = 3.85, p = .0002$], and this result was still significant when sex was included in the regression [$t(239) = 2.14, p = .0329$]. However, in a full analysis that included age as well as sex, the interaction of physician and age was significant, with the older laypeople less sensitive to probability and the older physicians no different from the younger physicians. No other interactions were significant, so we estimated the model with only the physician \times age interaction term. In this model, the main effects of physician and age were not significant, men were more

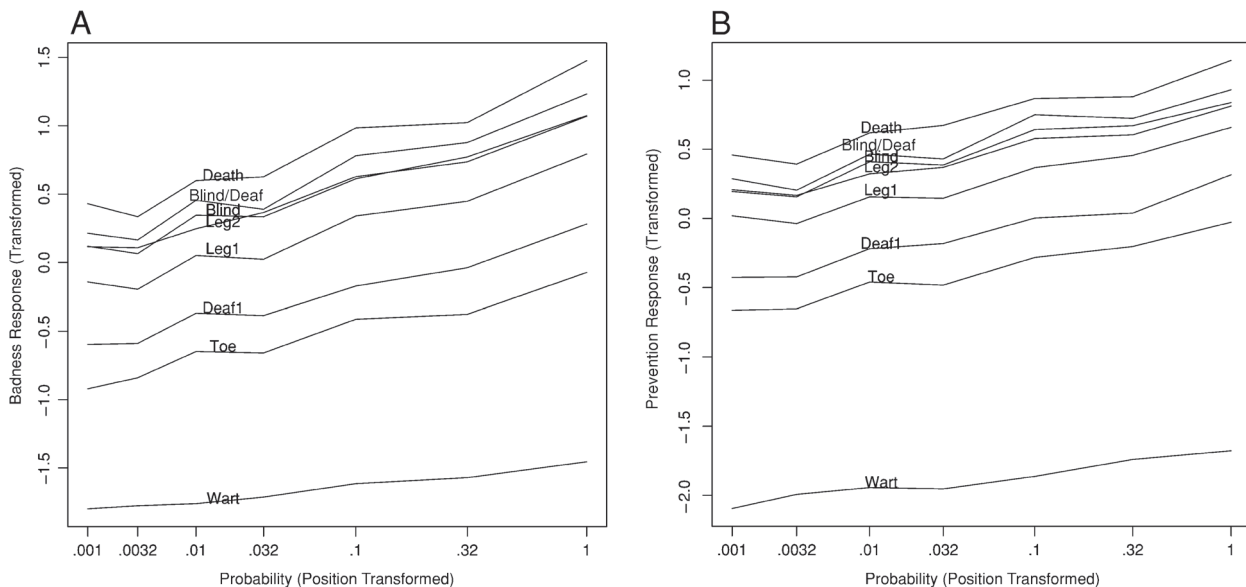


Figure 2. Transformed responses as a function of probability and condition.

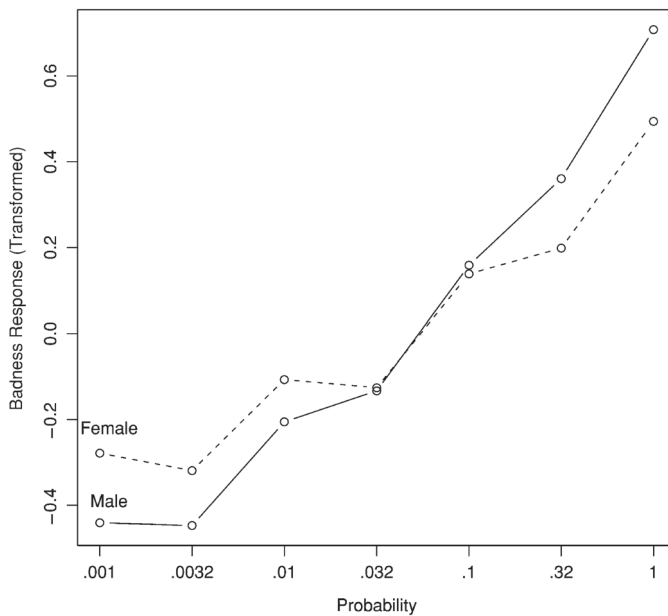


Figure 3. Mean probability weights for badness responses for males (solid line) and females (dashed line).

sensitive to probability than were women [difference of .089; $t(237) = 3.24, p = .0013$], and the physician \times age interaction was significant [$t(237) = 2.38, p = .0180$]. As has been noted, there was a strong effect of age for laypeople, a decline of .006 per year in the probability sensitivity measure [$t(106) = -3.20, p = .0018$, in a regression that included sex], and no significant effect for physicians (a decline per year of less than .001).

In sum, the physicians were more sensitive to probability than were the laypeople, but mostly at older ages. Men were, however, consistently more sensitive to probability than were women. Figure 3 shows the probability weights by sex.

We now will turn to the sample of those who made both badness and prevention judgments (which excluded the older, academic physicians). Because the probability sensitivities for the badness and prevention judgments did not differ in their predictors, we averaged them to form a single probability sensitivity measure. In a regression, physician was nonsignificant as a predictor of sensitivity, but sex, age, and speed were significant predictors, with greater sensitivity in males, younger subjects, and slower subjects [$t(163) = 3.70, p = .0003$, for age; $t(163) = 2.99, p = .0032$, for sex; $t(163) = 2.19, p = .0296$, for speed; no interactions were significant]. The negative results for physician are consistent with our finding with the full sample, where we inferred that the difference was mostly in older subjects.

Extreme Cases

So far, we have presented the formal analysis for probability and medical conditions. The EU model (with transformation of probability) fits the data from most subjects very well. But the model itself is extremely permissive, in

that it does not restrict the subject's implied values. Like Hume, it assumes that "It is not contrary to reason to prefer the destruction of the whole world to the scratching of my little finger." Yet some of the implied judgments seem like Hume's example.

Figure 4 shows the responses for the fifth of the subjects who had the lowest probability sensitivity to badness judgments (dashed lines) and the highest fifth (solid lines). The lowest group essentially ignores probability. They see a .001 chance of death as worse than a probability of 1 of any other condition (except being blind and deaf). In general, they consider a very low probability of a very serious condition as much worse than a higher probability of a less serious condition. The highest group attends much less to the nature of the condition, so they see the certainty of losing a toe as worse than a .032 chance of going blind and deaf. (If this example seems less extreme, perhaps, then, this group is distorting less, but we have no right answer.)

The variance of probability sensitivity was larger for the laypeople than for the physicians. The variance did not differ for the two sexes. The difference between the physicians and the laypeople was significant when the physicians (the new medical school graduates) were compared with the 54 laypeople less than 33 years old, whose mean age was close to (actually slightly less than) that of the physicians [variances, .051 and .024, respectively; variance ratio, 2.16; $F(49,58) = 2.16, p = .0053$].

DISCUSSION

The most striking finding is the existence of wide individual differences in sensitivity to probability, relative to severity or badness of conditions, in simple prospects

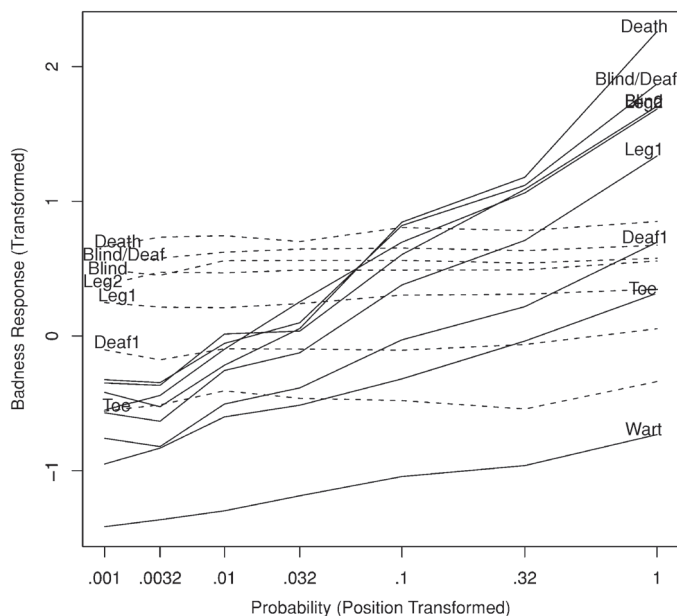


Figure 4. Transformed responses for subjects with lowest (dashed lines) and highest (solid lines) probability sensitivity.

consisting of a medical condition and a probability of getting it. For some subjects, the rankings of badness and of prevention importance were largely unaffected by probability. Few, if any, physicians were in this category. Thus, the laypeople's sensitivity to probability was more variable and generally lower than that of the physicians.

On the whole, however, differences between the physicians and the laypeople were limited to the older subjects in both groups. Sensitivity to probability decreased with age among the laypeople. Possibly, this age effect is the result of increased emphasis on probability in the school curriculum. The physicians did not show these age differences—presumably, either because they use probability more than does the average layperson in their work or because they were better educated about probability.

We also found consistent sex differences. Women were less sensitive to probability. These findings agree with other evidence indicating that females are less inclined to attend to probabilities (Gal & Baron, 1996).

If a physician is male and his patient female or if the patient is older, the patient may be more likely to view the risk of a low-probability outcome as worse than the physician views it. A physician can think of a “.001 chance of death” from a medical procedure as reassurance, but a patient who largely ignores probability will consider the procedure quite dangerous.

These differences between physicians and patients may help explain why some patients are generally dissatisfied with their communications with their physicians or why many patients feel that their physician minimizes their risks (Gurmankin, Baron, & Armstrong, 2004; Gurmankin et al., 2005). In addition, the insensitivity to differences in probability seen among some laypeople may help explain why some patients engage in high-risk behavior, such as

smoking, but will avoid lower risk activities, such as flying in a plane.

Our study was not designed to test conformity to EU theory systematically. Moreover, the judgments we obtained need not conform to EU unless they are components of decision making. Still, we found a couple of results indicating departures from EU.

First, prevention judgments were less affected by probability than were badness judgments. We thus failed to find *scale convergence* in the sense of Birnbaum and Veit (1974). Scale convergence implies that people have a single internal scale for badness of risks, although they may transform this scale into judgments by different transformations, depending on the task. It is possible that some people really do tend to ignore probability more when they think about prevention. In general, this result implies that at least one of the judgments does not order the prospects in the same way that EU would order them.

Second, as can be found in the evidence supporting prospect theory (Kahneman & Tversky, 1979, 2000), subjects are more affected by differences among high probabilities than by those among small ones. If judgments were ranked according to EU, the multiplication of probability by utility would be, in our analysis, equivalent to addition of the logarithms of probability and utility. Because our probabilities were logarithmically spaced, each step on our probability scale would have had the same effect on the transformed judgments. Yet we found that the probability transformation was convex—that is, the steps at the high end of the scale had relatively more influence than EU would predict.

A second possible interpretation of this convexity is that the subjects were, in effect, adding subjective probability and utility, rather than multiplying them. The steps

in our probability scale were larger steps for higher probabilities, in terms of probability differences, although not in terms of ratios (or therefore, in terms of logarithms). Thus, the convexity we found would be predicted if the subjects sometimes followed an additive model, instead of a multiplicative model (as was found by DeKay & McClelland, 1996, and Mellers & Chang, 1994), and if their subjective probabilities were a linear function of the stated probabilities.

One way in which subjects might make these judgments is anchoring and adjustment (as was found by Schkade & Johnson, 1989). Subjects would use the nature of the condition, which may be more salient and important to them, to make an initial rating and then adjust their response downward for lower probabilities. Such a two-step process would be consistent with the addition interpretation in the last paragraph, if the amount of adjustment for a given change in probability did not depend on the health condition. But it is also consistent with the prospect theory interpretation, if adjustment did depend on the starting point, as well as on the probability.

If such a two-step process occurs, we might expect some subjects to stop too soon, after insufficient adjustment. Such premature stopping would explain our finding that faster subjects were less sensitive to probability. Premature stopping would not explain the sex differences, however, since women were no faster than men, on average (and were nonsignificantly slower when age and physician were included in the regression). Sex differences might result from differences in the magnitude of the adjustment made, given that some adjustment was attempted.

Premature stopping could result either from simple impatience or from failure to recognize the relevance of probability information. Some people do not recognize the value or relevance of quantitative risk information to decisions involving risk (Baron, Granato, Spranca, & Teubal, 1993; Hogarth & Kunreuther, 1995; Huber, Wider, & Huber, 1997). In one study, for example, when subjects were required to seek out their own information to make hypothetical risky decisions, 78% did not ask for data on outcome probabilities (Huber et al., 1997). In an examination of people's decisions to purchase a warranty against failure of electronic products, subjects rarely mentioned the probability of need for repair as a factor in the decision to purchase the warranty (Hogarth & Kunreuther, 1995).

More work is needed to explore the effect of response to probability on physician-patient communication and patients' behaviors and medical decisions. Understanding individual differences and physician-patient differences may reduce problems in the physician-patient relationship and patient decision making by helping physicians to make accommodations for them in their communications with patients and by revealing points of intervention to optimize patient decision making and to increase patient preventive health behavior against higher probability outcomes.

Additional work is also needed to understand the mechanisms behind insensitivity to probability and those behind the individual differences. It would be of interest to see whether more direct evidence of an adjustment

process can be found (such as with the use of process tracing, as was done by Schkade & Johnson, 1989) and, if so, how explicit time pressure may affect it. Some medical decisions are made under time pressure (such as those in emergency rooms), and in these cases, people may be relatively insensitive to the probability of the consequences that they try to prevent.

Also of interest is whether insensitivity is still found in matching tasks that require attention to all dimensions. If insensitivity to probability is the result of a prominence effect, in which the outcome is more prominent (Tversky, Sattath, & Slovic, 1988), the use of a matching task should increase sensitivity to probability.

REFERENCES

- BARON, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 19-36). London: Blackwell.
- BARON, J., GRANATO, L., SPRANCA, M., & TEUBAL, E. (1993). Decision-making biases in children and early adolescents: Exploratory studies. *Merrill Palmer Quarterly*, **39**, 23-47.
- BIRNBAUM, M. H., & VEIT, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, **15**, 7-15.
- COLLINS, T., CLARK, J., PETERSEN, L., & KRESSIN, N. (2002). Racial differences in how patients perceive physician communication regarding cardiac testing. *Medical Care*, **40**(Suppl.), 127-134.
- DEKAY, M. L., & MCCLELLAND, G. H. (1996). Probability and utility components of endangered species preservation programs. *Journal of Experimental Psychology: Applied*, **2**, 60-83.
- GAL, I., & BARON, J. (1996). Understanding repeated choices. *Thinking & Reasoning*, **2**, 81-98.
- GURMANKIN, A. D., BARON, J., & ARMSTRONG, K. (2004). Intended message versus message received in hypothetical physician risk communication: Exploring the gap. *Risk Analysis*, **24**, 1337-1347.
- GURMANKIN, A. D., DOMCHEK, S., STOPFER, J., FELLS, C., & ARMSTRONG, K. (2005). Patients' resistance to risk information in genetic counseling for BRCA1/2. *Archives of Internal Medicine*, **165**, 523-529.
- HOGARTH, R. M., & KUNREUTHER, H. (1995). Decision making under ignorance: Arguing with yourself. *Journal of Risk & Uncertainty*, **10**, 15-36.
- HUBER, O., WIDER, R., & HUBER, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, **95**, 15-29.
- KAHNEMAN, D., & TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263-291.
- KAHNEMAN, D., & TVERSKY, A. (2000). *Choices, values, and frames*. New York: Cambridge University Press.
- KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- MELLERS, B. A., & CHANG, S. (1994). Representations of risk judgments. *Organizational Behavior & Human Decision Processes*, **57**, 167-184.
- NISBETT, R. E., FONG, G. T., LEHMAN, D. R., & CHENG, P. W. (1987). Teaching reasoning. *Science*, **238**, 625-631.
- R DEVELOPMENT CORE TEAM (2004). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at www.R-project.org.
- SCHKADE, D. A., & JOHNSON, E. (1989). Cognitive processes in preference reversals. *Organizational Behavior & Human Performance*, **44**, 203-231.
- TVERSKY, A. (1967). Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, **4**, 175-202.
- TVERSKY, A., SATTATH, S., & SLOVIC, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, **95**, 371-384.
- YAMAGISHI, K. (1997). When a 12.86% mortality is more dangerous than 21.14%: Implications for risk communication. *Applied Cognitive Psychology*, **11**, 495-506.

NOTES

1. The study was done in June 2003, immediately after completion of the last term of medical school. With the help of students at each school, we sent e-mail to the 4th-year classes of the University of Pennsylvania (150 students on the list), Cornell (105), Jefferson (180), Yale (165), Johns Hopkins (100), and Hahneman (230).

2. Note, in Figure 1, that badness and prevention judgments for medical conditions at probabilities of .001, .01, .1, and 1 were unusually high. These are the ones with powers of 10 and smaller denominators. It seems that the subjects were attending more to the denominator (the opposite of the effect observed by Yamagishi, 1997). Using the raw badness and prevention ratings, we carried out a separate analysis on the judgments of conditions at probabilities that were powers of 10 versus those that were not. We measured this denominator effect as the difference in the

response measure between the power-of-10 probabilities and the others. The denominator effect was larger when the probability slope—the raw regression weight for probability—was larger. In the limit, the subjects whose judgments were insensitive to probability could not show the effect. Thus, we included the probability slope in the regression. When this was done, we found that the physicians showed a smaller denominator effect than did the laypeople [$t(237) = 2.90, p = .0040$], with no other variables significant, except of course for probability slope ($t = 5.75, p = .0000$). We did not find a significant denominator effect for prevention judgments.

(Manuscript received June 24, 2004;
revision accepted for publication December 19, 2004.)