

The Psychology of Decision Making

Probability Judgment in Medicine:

Discounting Unspecified Possibilities

DONALD A. REDELMEIER, MD, DEREK J. KOEHLER, PhD,
VARDA LIBERMAN, PhD, AMOS TVERSKY, PhD

Research in cognitive psychology has indicated that alternative descriptions of the same event can give rise to different probability judgments. This observation has led to the development of a descriptive account, called support theory, which assumes that the judged probability of an explicit description of an event (that lists specific possibilities) generally exceeds the judged probability of an implicit description of the same event (that does not mention specific possibilities). To investigate this assumption in medical judgment, the authors presented physicians with brief clinical scenarios describing individual patients and elicited diagnostic and prognostic probability judgments. The results showed that the physicians tended to discount unspecified possibilities, as predicted by support theory. The authors suggest that an awareness of the discrepancy between intuitive judgments and the laws of chance may provide opportunities for improving medical decision making. *Key words*: probability judgment; support theory; unpacking principle; cognition. (*Med Decis Making* 1995;15:227-230)

Medical decisions are often made under uncertainty. When evaluating a patient with chest pain, for example, a physician needs to consider the possibility that the patient is having a myocardial infarction, the risk of a serious hemorrhage if thrombolytics are administered, and the consequences if thrombolytics are not administered. Uncertainty can sometimes be reduced by collecting additional data, reviewing the scientific literature, and consulting experts. However, it cannot always be eliminated in a timely manner.¹ As a consequence, action often depends on intuitive judgments of the likelihoods of various possibilities.

Research on judgment under uncertainty has shown

that both laypeople and experts do not always follow the principles of probability theory.^{2,3} In particular, alternative representations of the same possibility can give rise to different probability judgments.⁴ To account for such observations, Tversky and Koehler⁵ have developed an account in which probability is assigned not to events-as in other models-but rather to descriptions of events, called *hypotheses*. This account, called *support theory*, assumes that each hypothesis refers to a unique event, but that a given event can be described by more than one hypothesis. For example, the explicit hypothesis "death due to traffic accident, drowning, electrocution, or any other accident" and the implicit hypothesis "death due to an accident" represent different descriptions of the same event.

The central assumption of support theory is the *unpacking principle*: providing a more detailed description of an implicit hypothesis generally increases its judged probability. Thus, the judged probability of the explicit description that lists various accidents generally exceeds the judged probability of the implicit description that does not mention specific accidents. Like the measured length of a coastline, which increases as the map becomes more detailed, the perceived likelihood of an event increases as its description becomes more specific. Both memory and attention contribute to this effect: unpacking can remind people of possibilities they might have overlooked, and the explicit mentioning of a possibility may increase

Received January 14, 1994, from the University of Toronto, Toronto, Ontario, Canada (DAR); the Wellesley Hospital Research Institute, Toronto (DAR); Stanford University, Stanford, California (DJK, AT); and the Open University of Israel, Tel Aviv, Israel (VL). Revision accepted for publication August 18, 1994. Supported by a grant from the PSI Foundation of Ontario, by Grant #SBR-9408684 from the National Science Foundation, and by a Fulbright Visiting Scholarship Award. Dr. Redelmeier was supported by a career scientist award from the Ontario Ministry of Health, Dr. Koehler by a National Defence Science and Engineering Graduate Fellowship, and Drs. Liberman and Tversky by grant #92-00389 from the United States-Israel Binational Science Foundation.

Address correspondence and reprint requests to Dr. Redelmeier: Clinical Epidemiology Division, The Wellesley Hospital Research Institute, 160 Wellesley St., East, 650 Turner Wing, Toronto, Ontario, Canada M4Y 1J3.

its salience and hence its perceived likelihood.

In accord with the classic theory of probability, support theory assumes that the judged probability of a hypothesis and of its complement add to unity. For example, the judged probability of the hypothesis “death due to a natural cause” and that of the hypothesis “death due to an unnatural cause” should sum to one, even though each judgment could be increased by unpacking the respective category. The unpacking effect, as well as binary complementarity, have been observed in several experiments involving nonmedical situations.⁵ The present article explores these principles in medical judgments. To do so, we presented clinicians with brief scenarios describing an individual patient and asked them to judge the probabilities of relevant medical possibilities.

Unpacking the Residual

In a survey of house officers ($n = 59$) at Stanford University, physicians were asked to review the following scenario:

A well-known 22-year-old Hollywood actress presents to the emergency department with pain in the right lower quadrant of her abdomen of 12 hours' duration. Her last normal menstrual period was four weeks ago.

Half the physicians, selected at random, were asked to estimate probabilities for two diagnoses (“gastroenteritis” and “ectopic pregnancy”) and the residual category (“none of the above”). The other physicians were asked to estimate probabilities for the following five diagnoses: the two diagnoses specified above (“gastroenteritis” and “ectopic pregnancy”), three additional specific diagnoses (“appendicitis,” “pyelonephritis,” and “pelvic inflammatory disease”), and the residual category (“none of the above”). The two tasks differed only in that the residual category in the first (short) list was partially unpacked in the second (long) list. All the physicians were told that the patient had only one condition and, hence, that the judged probabilities should add to 100%.

Logically, the probability of the residual “none of the above” in the short list should equal the sum of the probabilities of the corresponding possibilities in the long list. In accord with the unpacking principle, however, we found that the average probability assigned to the residual in the short list was smaller than the sum of the corresponding probabilities in the long list (50% vs 69%, $p < 0.005$ by Mann-Whitney test). As a consequence, unpacking the residual category changed the probabilities assigned to specific diagnoses. For example, the average probability assigned to “gastroenteritis” was substantially higher in the short list than in the long list (31% vs 16%, $p < 0.005$ by Mann-Whitney test). Evidently, unpacking

the residual hypothesis reminded physicians of diseases they might have overlooked, or increased the salience of diagnoses that they had considered.

Highlighting One Possibility

In the previous example physicians were asked to assign probabilities to a set of possibilities. Often, however, physicians focus on a single possibility. In this case, they may be prone to overestimate the likelihood of that possibility because its alternatives are unspecified. To illustrate this point, we presented the following scenario to a group of expert physicians ($n = 52$) at Tel Aviv University:

R.G. is a 67-year-old retired farmer who presents to the emergency department with chest pain of four hours' duration. The diagnosis is acute myocardial infarction. Physical examination shows no evidence of pulmonary edema, hypotension, or mental status changes. His EKG shows ST-segment elevation in the anterior leads, but no dysrhythmia or heart block. His past medical history is unremarkable. He is admitted to the hospital and treated in the usual manner. Consider the possible outcomes.

Each physician was randomly assigned to evaluate one of the following four prognoses for this patient: “dying during this admission,” “surviving this admission but dying within one year,” “living for more than one year but less than ten years,” or “surviving for more than ten years.” The average probabilities assigned to these prognoses were 14%, 26%, 55%, and 69%, respectively. According to standard theory, the probabilities assigned to these outcomes should sum to 100%. In contrast, the average judgments added to 164% (95% confidence interval: 134% to 194%). As implied by the unpacking principle, the physicians in each group overweighted the possibility that was explicitly mentioned relative to the unspecified alternative. All groups, indeed, overestimated the frequencies reported in the literature.⁶ Notice that while the results of the previous problem can be interpreted as a memory effect (reminding physicians of additional possibilities), the present results represent an attention effect (highlighting a particular interval on a continuum).

Binary Complementarity

We have attributed the preceding results to the unpacking principle. An alternative interpretation is that people overestimate the (focal) hypothesis that they are asked to evaluate. If this interpretation be correct, the sum of the judged probabilities for a pair of complementary hypotheses should exceed one. To test

this prediction, we presented the preceding scenario to fourth-year medical students ($n = 149$) at the University of Toronto. Half the participants, selected at random, were asked to evaluate the probability that the patient would "survive this hospitalization." The other half were asked to evaluate the probability that the patient would "die during this hospitalization." We found that the mean judged probabilities in the two groups were 78% and 21% respectively, summing to 99% (95% confidence interval: 94% to 104%). As implied by support theory, judged probabilities add to 100% in cases with only two possibilities, and exceed 100% in cases involving more than two possibilities. This observation demonstrates that people overestimate what is specified, not what is under evaluation.

Treatment Decisions

The final example shows that the unpacking effect is not limited to probability judgments but can also extend to treatment decisions. We asked fourth-year medical students ($n = 148$) at the University of Toronto to consider the following scenario:

MS. is a 42-year-old journalist who presents to the emergency department because of a fever and headache of two days' duration. Past medical history is remarkable only for 15 years of lupus erythematosus, controlled on Tylenol and chronic steroids (prednisone 10 mg daily). She does not look sick. Vital signs are normal. Physical examination reveals tenderness over the frontal sinuses and pharyngeal erythema. There is no neck stiffness, tympanic membrane redness, or cervical adenopathy. The remainder of the physical examination is unremarkable aside from some degenerative changes in the small joints of both hands.

For half the students, selected at random, the scenario was followed by the sentence: "Obviously, many diagnoses are possible given this limited information, including CNS vasculitis, lupus cerebritis, intracranial opportunistic infection, sinusitis, and a subdural hematoma." The other half were presented with a shorter sentence: "Obviously, many diagnoses are possible given this limited information, including sinusitis." Individuals in both groups were asked to indicate whether they would recommend ordering a CAT scan of the head.

Logically, there should be no difference between the responses to the two versions because both describe the same situation. On the other hand, support theory suggests that the possibility of sinusitis will loom larger when it is the only specified diagnosis than when it is accompanied by other specified diagnoses. Consequently, we expected that fewer physicians would order a CAT scan in the short version because the diagnosis of sinusitis does not normally call for this

test.' Indeed, we found that fewer respondents recommended a CAT scan in response to the short version than in response to the long version (20% vs 32%) ($p < 0.05$ by Mann-Whitney test). Thus, the unpacking principle applies to treatment recommendations, not only to probability judgments.

Conclusion

Subjective assessments of uncertain events are sometimes necessary, even though they are often fallible. In this study we focused on a particularly significant source of error, namely, the tendency to discount unspecified possibilities. In the first problem we demonstrated the unpacking effect in a diagnostic task by reminding physicians of possibilities they might have overlooked. In the second problem we obtained the same effect in a prognostic task by highlighting a specific interval along a continuum. In the third problem we showed that the unpacking effect cannot be explained by overestimating the focal possibility. And in the final problem we illustrated the unpacking effect in a decision task. Together, the findings confirm the main qualitative predictions of support theory in medical judgment.

It could be argued that our respondents believed that the request to evaluate a particular hypothesis conveys relevant information and suggests that the hypothesis in question is not improbable. Although such belief may contribute to the unpacking effect, it does not fully explain the data. First, this account implies an overweighting of the focal hypothesis, contrary to the finding of binary complementarity. Second, the unpacking effect was pronounced in the myocardial infarction example, where the experts were informed that other physicians were evaluating different hypotheses. Finally, the unpacking effect has also been observed in non-medical problems where the respondents were made aware that the focal hypothesis had been randomly chosen.⁵

Although there is no simple method for eliminating the unpacking effect, we call attention to its presence and suggest some corrective procedures. First, clinicians need to recognize that judgments under uncertainty are susceptible to error; in particular, alternative descriptions of the same situation may lead to different judgments. Second, clinicians should be encouraged to unpack broad categories and compare possibilities at similar levels of specificity, rather than compare a single specific possibility against an unspecified set of alternatives. Indeed, unpacking the implicit complement of a focal hypothesis may serve as a useful method for reducing overconfidence. More generally, a better understanding of the cognitive psychology underlying medical judgment could help identify common biases and suggest corrective procedures.

References

1. Pauker SG, Kopelman RI. How sure is sure enough? *N Engl J Med.* 1992;326:688-91.
2. Kahneman D, Slovic P, Tversky A (eds). *Judgment under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press, 1982.
3. von Winterfeldt D, Edwards W. *Decision Analysis and Behavioral Research.* New York: Cambridge University Press, 1986.
4. Fischhoff B, Slovic P, Lichtenstein S. *Fault trees: sensitivity of estimated failure probabilities to problem representation.* *J Exp Psychol Hum Percept Perform.* 1978;3:552-64.
5. Tversky A, Koehler DJ. Support theory: a nonextensional representation of subjective probability. *Psychol Rev.* 1994;101:547-67.
6. Goldberg RJ, Gore JM, Alpert JS, et al. Cardiogenic shock after acute myocardial infarction: incidence and mortality from a community-wide perspective 1975 to 1988. *N Engl J Med.* 1991;325:1117-22.
7. Hourihane J. CT scans and the common cold. *N Engl J Med.* 1994;330:1826-7.

Correction

An error appeared in the article, "Global judgments versus decision-Model-facilitated judgments," *Med Decis Making* 1994;14:19-26, by Oddone et al. In the original Table 2, the authors reported estimates of operative morbidity and mortality for a specific clinical scenario. The values used in their calculations of quality-adjusted life expectancy and reported in the results section were based on estimates for operative morbidity and mortality each of which was at the midpoint of a range of estimates that the expert panel agreed upon in a group discussion. The authors believed that these values were more representative than mean values. However, these values were not used in the decision model. Corrected values for Table 2 are shown here. Because one value was used for either high or low surgical risk, there is no variation. The delta quality-adjusted life expectancies in the text and the figures are still correct.

Corrected values for Table 2.

	Operative Mortality	Operative Morbidity
Multiple episodes carotid TIA/AF 70-99%, low surgical risk	0.005	0.015
Post-atheroembolic stroke, 30-49% ipsilateral, high surgical risk	0.015	0.025
Asymptomatic, 70-99% ipsilateral, low surgical risk	0.005	0.015
Asymptomatic, preop CABG 50-69% ipsilateral, high surgical risk	0.015	0.025