IMPROVING STATISTICAL REASONING BY USING THE RIGHT REPRESENTATIONAL FORMAT

Peter Sedlmeier
Chemnitz University of Technology
Germany

New results in research on judgment under uncertainty show a way of how to improve the teaching of statistical reasoning. The implications of this research are that (i) successful learning needs doing, and (ii) that the format in which information is represented plays a decisive role. Statistical problems are, for instance, solved much better if the relevant pieces of information are presented as frequencies rather than probabilities. It also helps a lot if random processes can be observed rather than only read about. A computer program is presented that incorporates these implications from psychological research. The software accompanies an elementary text book on probability theory to be used in high school.

INTRODUCTION

The TIMS (Third International Mathematics and Science) study has revealed that German as well as American high school students show severe deficits in many areas of mathematics, including probability theory (Baumert & Lehman, 1997). This is consistent with results from psychology research: Adults show poor performance in tasks that deal with judgment under uncertainty, that is, with probabilities (e.g. Kahneman, Slovic & Tversky, 1982). In the area of judgment and decision making the still-dominant view holds that difficulties in adequately dealing with uncertain information – sometimes termed “cognitive illusions” – are as resistant to training attempts as are visual illusions (Gigerenzer, 1991). However, increasing evidence shows that this need not remain so: If the tasks that yielded such poor results are changed only slightly, the solution rates go up considerably (e.g., Gigerenzer, Hertwig, Hoffrage & Sedlemeier, in press). Tasks have to be presented in a way that invokes pre-existing valid intuitions, that is, spontaneous tendencies to judge or to act in a certain way.

THE ROLE OF INTUITIONS

It can be shown that whether there is a match between a task and a corresponding intuition makes a big difference in whether things of daily use, such as stoves, doors, light switches or flight schedules, are easy to use or not (Norman, 1988). The use of intuitions has also been explored in mathematics education. Consider the following example from Fishbein (1994):

Problem 1: From 1 quintal of wheat, you get 0.75 quintals of flour. How much flour do you get from 15 quintals of wheat?

Problem 2: 1 kilo of a detergent is used in making 15 kilos of soap. How much soap can be made from 0.75 kilos of detergent?

It turns out that Problem 1 is much easier to solve for 5^{th}, 7^{th} and 9^{th} grade students than Problem 2, although mathematically the solutions are the same. The difference lies in whether one multiplies 0.75 by 15 as in Problem 1 or whether the multiplication is 15 times 0.75 as in Problem 2. Problem 1 conforms to the intuition that if a number is multiplied, it gets bigger whereas Problem 2 contradicts this intuition.

COMPUTER PROGRAM: SOME BACKGROUND INFORMATION

The computer program described here accompanies a textbook on probability theory for German high school students (Sedlmeier & Köhlers, 2001) and exploits just these pre-existing correspondences between valid intuitions and probability tasks. Large parts of the program have been tested successfully in empirical studies (Sedlmeier, 1999, 2000; Sedlmeier & Gigerenzer, 2001). The success of the training program was due to two ingredients: the use of suitable representational formats and learning by doing. Representational formats that are close to everyday life are more likely to evoke valid intuitions than are very abstract ones. In the case of probability theory, this means that if the information in probability tasks is given in frequencies, these tasks are much easier to solve than if the information is given in probabilities, which are quite new in human history; also, if one experiences random sampling or can perform it oneself, probability tasks are solved more easily than if one only reads about random samples (Sedlmeier, 1998). Learning by doing is realized in the program by allowing the users to manipulate graphical interfaces and solve tasks on their own, guided by the program's feedback.

ILLUSTRATIVE EXAMPLE

An example illustrates how the program works (for a more thorough description see Sedlmeier, 2001). Consider the following problem:

A reporter for a women's monthly magazine would like to write an article about breast cancer. As a part of her research, she focuses on mammography as an indicator of breast cancer. She wonders what it really means if a woman tests positive for breast cancer during her routine mammography examination. She has the following data:

- The probability that a woman who undergoes a mammography examination will have breast cancer is 1%
- If a woman undergoing a mammography has breast cancer, the probability that she will test positive is 80%.
- If a woman undergoing a mammography does not have breast cancer, the probability that she will test positive is 10%.

What is the probability that a woman who has undergone a mammography actually has breast cancer, if she tests positive?

This problem was solved correctly by less than 10% of participants, experts (medical doctors) and laypeople alike, in several studies. The usual way to solve it is to apply Bayes's formula [p(cancer) = 0.01, p(pos|cancer) = 0.8, and p(pos|no cancer) = 0.1, as given in the text; p(no cancer) can be calculated as 1 – p(cancer) = 0.99]:

$$\begin{aligned} p(cancer \mid pos) &= \frac{p(cancer) \times p(pos \mid cancer)}{p(cancer) \times p(pos \mid cancer) + p(no\ cancer) \times p(pos \mid no\ cancer)} \\ &= \frac{.01 \times .8}{.01 \times .8 + .99 \times .1} \\ &\approx 7.5\% \end{aligned}$$

REPRESENTATIONAL FORMATS

If participants are taught how to extract the relevant information from texts and how to use it in the formula, there is some training effect right after the training, but it is almost completely gone one week later. This way of solving a probability revision problem (the probability of a woman having breast cancer is *revised* in the light of a positive test result) does not seem to evoke any valid statistical intuitions.

How does the program deal with problems of this kind? It offers two representational formats, the frequency grid and the frequency tree. A frequency grid (not shown) could, for instance, consist of 1,000 squares representing a random sample of 1,000 women. Ten of these women (1%) can be expected to have breast cancer and the corresponding squares are colored. From these 10, 8 (80%) can expect a positive test result (corresponding squares marked by a cross) but from the remaining 990 women, also 99 (10%) can expect such a positive result (corresponding squares also marked by a cross). The sought for conditional probability of breast cancer given a positive test result is just the number of women who have cancer *and* are tested positive (number of squares that are both colored and marked by a cross) divided by all the women with a positive test result (total number of squares marked by a cross): 8 / 107 = 0.075. When participants received a frequency-grid training, the high immediate training effect (about 90% correct solutions in probability revision problems not used in the training) was found to remain stable in retests up to three months afterward.

Basically the same result was obtained with a different kind of frequency representation, the frequency tree (Figure 1). Again, a random sample of 1,000 women (top node in Figure 1) is divided up into women with ("Brustkrebs") and without ("Kein Brustkrebs") breast cancer (two middle nodes) and these are divided up into women with and without a positive test result (lower nodes). Again, it suffices to divide the result in the left lower node (number of women with breast cancer and positive test result) by the number of all women with a positive test result.

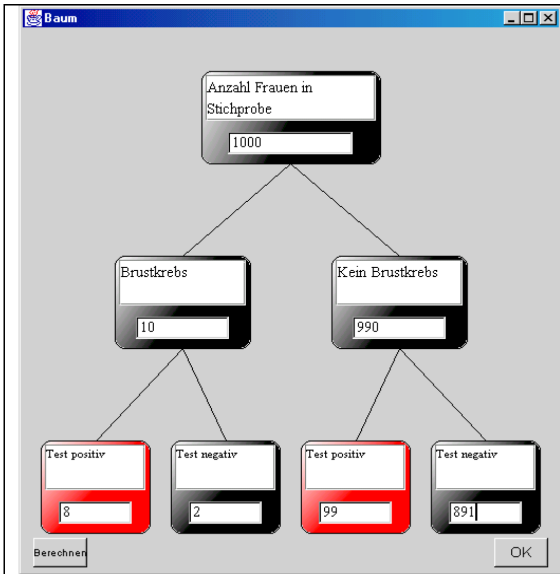

Figure 1. Frequency Tree.

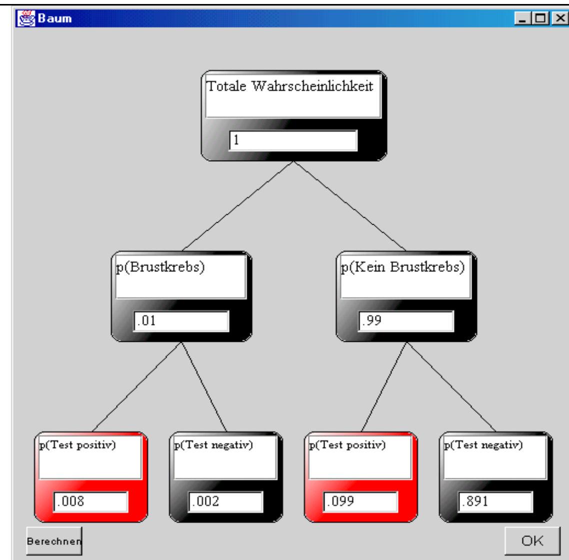

Figure 2. Probability Tree as Used in the Training Studies.

A POSSIBLE OBJECTIION

However, one could make the point that the formula training differs from the frequency training also in another respect: non-graphical vs. graphical representation. Do the graphics play a crucial role? This question was examined in a study that compared the frequency tree to a probability tree (Figure 2). The probability tree is equivalent to the frequency tree except that now numbers are confined to values between 0 and 1 – otherwise calculations remain the same. There is, however, a dramatic difference in training success between the two kinds of representational formats: Whereas both formats yielded comparable short-term training effects, the long-term effect for the probability tree was as low as that achieved with the formula (Sedlmeier & Gigerenzer, 2001).

OTHER FEATURES OF THE PROGRAM

The computer program also contains an animated frequency representation, the “virtual urn,” that makes random sampling visible. The virtual urn can be filled with all kinds of discrete (population) distributions with the events represented as balls marked with colors and symbols that can be chosen freely. The animated sampling procedure can be performed with and without replacement. The virtual urn is the basis for all simulations in the program. It could be shown empirically that the use of the virtual urn helps considerably in understanding and solving probability tasks (Sedlmeier, 1999). The computer program covers all aspects of basic probability theory treated in German high schools, including confidence intervals and significance tests for binomial distributions.

REFERENCES

Baumert, J., & Lehman, R. (1997). *TIMSS—Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen, Germany: Leske & Budrich.

Fischbein, E. (1994). The interaction between the formal, the algorithmic, and the intuitive components in a mathematical activity. In R. Biehler, R. W. Scholz, R. Strässer, and B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (pp. 231-245). Dordrecht: Kluwer

Gigerenzer, G. (1991). On cognitive illusions and rationality. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *21*, 225-249.

Gigerenzer, G., Hertwig, R., Hoffrage, U., & Sedlmeier, P. (in press). Cognitive illusions reconsidered. In C.R. Plott, and V.L. Smith, (Eds.), *Handbook of experimental economics results*. North Holland/Elsevier Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Norman, D.A. (1988). *The design of everyday things*. New York: Doubleday.

Sedlmeier, P. (1998). The distribution matters: Two types of sample-size tasks. *Journal of Behavioral Decision Making*, *11*, 281-301.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Erlbaum.

Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. *Instructional Science*, *28*, 227-262.

Sedlmeier, P. (2001). Statistik ohne Formeln. In M. Borovcnik, J. Engel, and D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht*., (pp. 83-95). Hildesheim: Franzbecker.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*, 380–400.

Sedlmeier, P., & Köhlers, D. (2001). *Wahrscheinlichkeiten im Alltag: Statistik ohne Formeln*. Braunschweig: Westermann.