Statistical Science
2005, Vol. 20, No. 3, 223–230
DOI 10.1214/088342305000000296

How to Confuse with Statistics or: The Use and Misuse of Conditional Probabilities

Walter Krämer and Gerd Gigerenzer

Abstract. This article shows by various examples how consumers of statistical information may be confused when this information is presented in terms of conditional probabilities. It also shows how this confusion helps others to lie with statistics, and it suggests both confusion and lies can be exposed by using alternative modes of conveying statistical information.

Key words and phrases: Conditional probabilities, natural frequencies, heuristical reasoning.

1. INTRODUCTION

"The notion of conditional probability is a basic tool of probability theory, and it is unfortunate that its great simplicity is somehow obscured by a singularly clumsy terminology" (Feller, 1968, page 114). Below we argue that what Feller has rightly called a "singularly clumsy terminology," in addition to obscuring the basic simplicity of concepts and ideas, easily lends itself to intentional and unintentional misinterpretation of statistical information of many sorts. Examples in Darrel Huff's book are mainly in the chapter on semiattached figures, for instance, when discussing fatalities on highways on page 78: "Four times more fatalities occur on the highways at 7 p.m. than at 7 a.m." Huff points out that this of course does not imply, as some newspaper had suggested, that it is more dangerous to drive in the evening than in the morning. Recast in the language of conditional probabilities, what Huff observes is that $P(\text{accident} \mid \text{7 p.m.})$ should not be confused with $P(\text{7 p.m.} \mid \text{accident})$. Unfortunately, it was.

Although the term *conditional probability* does not appear once in Huff's remarkable book, it is clear that many other examples of statistical abuse that he discovered can be rephrased in terms of conditional probabilities. Below we survey various ways in which such reasoning can be misleading, and we provide some fresh examples. We also show that the potential for confusion is easily reduced by abandoning the conventional, "singularly clumsy terminology" of conditional probabilities in favor of presentation of information in terms of natural frequencies.

Walter Krämer is Professor of Business and Economic Statistics, Fachbereich Statistik, Universität Dortmund, 44221 Dortmund, Germany (e-mail: walterk@statistik.uni-dortmund.de). Gerd Gigerenzer is Director of the Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany (e-mail: gigerenzer@mpib-berlin.mpg.de).

2. FALLACIES IN ENUMERATION

One class of errors involving conditional probabilities comprises outright mistakes in computing them in the first place. One instance of consciously exploiting such computational errors in order to cheat the public is a game of cards called "Three Cards in a Hat," which used to be offered to innocent passers-by at country fairs in Germany and elsewhere. One card is red on both sides, one is white on both sides, and the other is red on one side and white on the other. The cheat draws one card blindly, and shows, for example, a red face up. The cheat then offers a wager of 10 Deutschmarks that the hidden side is also red.

The passer-by is assumed to argue like this: "The card is not the white-white one. Therefore, its hidden side is either red or white. As both cases are equally likely, the probability that the hidden side of the card on the table is red is $1/2$, so the wager is fair and can be accepted."

In fact, of course, the red-red card has probability $2/3$, since it can be drawn in two equally probable ways (one face up or the other face up, each of which will display red). The example therefore boils down to an incorrect enumeration of simple events in a Laplace experiment in the subpopulation composed of the remaining possibilities. As such, it has famous antecedents: The erroneous assignment by d'Alembert (1779, entry "Croix ou pile") of a probability of $1/3$ for heads-heads when twice throwing a coin, or the equally erroneous assertion by Leibniz (in a letter to L. Bourgnet from March 2, 1714, reprinted in Leibniz, 1887, pages 569–570) that, when throwing two dice, a sum of 11 is as likely as a sum of 12. A sum of 11, so he argued, can be obtained by adding 5 and 6, and a sum of 12 by adding 6 and 6. It did not occur to him that there are two equally probable ways of adding 5 and 6, but only one way to obtain 6 and 6.

Given illustrious precedents such as these, it comes as no surprise that wrongly inferred conditional and unconditional probabilities are lurking everywhere. Prominent textbook examples are the paradox of the second ace or the problem of the second boy (see, e.g., Bar-Hillel and Falk, 1982), not to mention the famous car-and-goat puzzle, also called the Monty Hall problem, which has engendered an enormous literature of its own. These puzzles are mainly of interest as mathematical curiosities and they are rarely used for statistical manipulation. We shall not dwell on them in detail here, but they serve to point out what many consumers of statistical information are ill-prepared to master.

3. CONFUSING CONDITIONAL AND CONDITIONING EVENTS

German medical doctors with an average of 14 years of professional experience were asked to imagine using a certain test to screen for colorectal cancer. The prevalence of this type of cancer was 0.3%, the sensitivity of the test (the conditional probability of detecting cancer when there is one) was 50% and the false positive rate was 3% (Gigerenzer, 2002; Gigerenzer and Edwards, 2003). The doctors were asked: "What is the probability that someone who tests positive actually has colorectal cancer?" The correct answer is about 5%. However, the doctors' answers ranged from 1% to 99%, with about half of them estimating this probability as 50% (the sensitivity) or 47% (the sensitivity minus the false positive rate).

The most common fault was to confuse the conditional probability of cancer, given that the test is positive, with the conditional probability that the test is positive, given that the individual has cancer. An analogous error also occurs when people are asked to interpret the result of a statistical test of significance, and sometimes there are disastrous consequences. In the fall of 1973 in the German city of Wuppertal, a local workman was accused of having murdered another local workman's wife. A forensic expert (correctly) computed a probability of only 0.027 that blood found on the defendant's clothes and on the scene of the crime by chance matched the victim's and defendant's blood groups, respectively. From this figure the expert then derived a probability of 97.3% for the defendant's guilt, and later, this probability came close to 100% by adding evidence from textile fibers. Only a perfect alibi saved the workman from an otherwise certain conviction (see the account in Ziegler, 1974).

Episodes such as this have undoubtedly happened in many courtrooms all over the world (Gigerenzer, 2002). On a formal level, a probability of 2.7% for the observed data, given innocence, was confused with a probability of 2.7% for innocence, given the observed data. Even in a Bayesian setting with certain a priori probabilities for guilt and innocence, one finds that a probability of 2.7% for the observed data given innocence does not necessarily translate into a probability of 97.3% that the defendant is guilty. And from the frequentist perspective, which is more common in forensic science, it is nonsense to assign a probability to either the null or to the alternative hypothesis.

Still, students and, remarkably, teachers of statistics, often misread the meaning of a statistical test of significance. Haller and Krauss (2002) asked 30 statistics instructors, 44 statistics students and 39 scientific psychologists from six psychology departments in Germany about the meaning of a significant two-sample t-test (significance level $= 1\%$). The test was supposed to detect a possible treatment effect based on a control group and a treatment group. The subjects were asked to comment upon the following six statements (all of which are false). They were told in advance that several or perhaps none of the statements were correct.

(1) You have absolutely disproved the null hypothesis that there is no difference between the population means. ○ true/false ○
(2) You have found the probability of the null hypothesis being true. ○ true/false ○
(3) You have absolutely proved your experimental hypothesis that there is a difference between the population means. ○ true/false ○
(4) You can deduce the probability of the experimental hypothesis being true. ○ true/false ○

(5) You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision. ○ true/false ○

(6) You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions. ○ true/false ○

All of the statistics students marked at least one of the above faulty statements as correct. And, quite disconcertingly, 90% of the academic psychologists and 80% of the methodology instructors did as well! In particular, one third of both the instructors and the academic psychologists and 59% of the statistics students marked item 4 as correct; that is, they believe that, given a rejection of the null at level 1%, they can deduce a probability of 99% that the alternative is correct.

Ironically, one finds that this misconception is perpetuated in many textbooks. Examples from the American market include Guilford (1942, and later editions), which was probably the most widely read textbook in the 1940s and 1950s, Miller and Buckhout (1973, statistical appendix by Brown, page 523) or Nunnally (1975, pages 194–196). Additional examples are collected in Gigerenzer (2000, Chapter 13) and Nickerson (2000). On the German market, there is Wyss (1991, page 547) or Schuchard-Fischer et al. (1982), who on page 83 of their best-selling textbook explicitly advise their readers that a rejection of the null at 5% implies a probability of 95% that the alternative is correct.

In one sense, this error can be seen as a probabilistic variant of a classic rule of logic (modus tollens): (1) “All human beings will eventually die” and (2) “Socrates is a human being” implies (3) “Socrates will die.” Now, what if (1) is not necessarily true, only highly probable [in the sense that the statement “If $A(=\text{human being})$ then $B(=\text{eventual death})$” holds not always, only most of the time]? Does this imply that its logical equivalent “If not B then not A” has the same large probability attached to it? This question has led to a lively exchange of letters in *Nature* (see Beck-Bornholdt and Dubben, 1996, 1997; or Edwards, 1996), which culminated in the scientific proof that the Pope is an alien: (1) A randomly selected human being is most probably not the Pope (the probability of selecting the Pope is $1:6$ billion $= 0.000\,000\,000\,17$). (2) John Paul II is the Pope. (3) Therefore, John Paul II is most probably not a human being.

Setting aside the fact that John Paul II has not been randomly selected from among all human beings, one finds that this argument again reflects the confusion that results from interchanging conditioning and conditional events. It is based on taking as equal the conditional probabilities $P(\text{not Pope} \mid \text{human})$ and $P(\text{not human} \mid \text{Pope})$. Since

$$P(\bar{A} \mid B) = P(\bar{B} \mid A) \quad \Longleftrightarrow \quad P(A \mid B) = P(B \mid A),$$

this is equivalent to taking as equal, in a universe comprised of humans and aliens, the conditional probabilities $P(\text{Pope} \mid \text{human})$ and $P(\text{human} \mid \text{Pope})$, which is nonsense. Or in terms of rules of logic: If the statement “If human then not Pope” holds most of the time, one cannot infer, but sometimes does, that its logical equivalent “If Pope then not human” likewise holds most of the time.

Strange as it may seem, this form of reasoning has even made its way into the pages of respectable journals. For instance, it was used by Leslie (1992) to prove that doom is near (the “doomsday argument”; see also Schrage, 1993). In this case the argument went: (1) If mankind is going to survive for a long time, then all human beings born so far, including myself, are only a small proportion of all human beings that will ever be born (i.e., the probability that I observe myself is negligible). (2) I observe myself. (3) Therefore, the end is near.

This argument is likewise based on interchanging conditioning and conditional events. While it is perfectly true that the conditional probability that a randomly selected human being (from among all human beings that have ever been and will ever be born) happens to be me, given that doom is near, is much larger that the conditional probability of the same event, given that doom is far away, one cannot infer from this inequality that the conditional probability that doom is near, given my existence, is likewise much larger than the conditional probability that doom is far away, given my existence. More formally: while the inequality in the following expression is correct, the equality signs are not:

$$\begin{aligned} P(\text{doom is near} \mid \text{me}) &= P(\text{me} \mid \text{doom is near}) \\ &\gg P(\text{me} \mid \text{doom far away}) \\ &= P(\text{doom far away} \mid \text{me}). \end{aligned}$$

4. CONDITIONAL PROBABILITIES AND FAVORABLE EVENTS

The tendency to confuse conditioning and conditional events can also lead to other incorrect conclusions. The most popular one is to infer from a conditional probability $P(A \mid B)$ that is seen as “large” that the conditional event A is "favorable" to the conditioning event B. This term was suggested by Chung (1942) and means that

$$P(B \mid A) > P(B).$$

This confusion occurs in various contexts and is possibly the most frequent logical error that is found in the interpretation of statistical information. Here are some examples from the German press (with the headlines translated into English):

- "Beware of German tourists" (according to *Der Spiegel* magazine, most foreign skiers involved in accidents in a Swiss skiing resort came from Germany).
- "Boys more at risk on bicycles" (the newspaper *Hannoversche Allgemeine Zeitung* reported that among children involved in bicycle accidents the majority were boys).
- "Soccer most dangerous sport" (the weekly magazine *Stern* commenting on a survey of accidents in sports).
- "Private homes as danger spots" (the newspaper *Die Welt* musing about the fact that a third of all fatal accidents in Germany occur in private homes).
- "German shepherd most dangerous dog around" (the newspaper *Ruhr-Nachrichten* on a statistic according to which German shepherds account for a record 31% of all reported attacks by dogs).
- "Women more disoriented drivers" (the newspaper *Bild* commenting on the fact that among cars that were found entering a one-way-street in the wrong direction, most were driven by women).

These examples can easily be extended. Most of them result from unintentionally misreading the statistical evidence. When there are cherished stereotypes to conserve, such as the German tourist bullying his fellow vacationers, or women somehow lost in space, perhaps some intentional neglect of logic may have played a role as well. Also, not all of the above statements are necessarily false. It might, for instance, well be true that when 1000 men and 1000 women drivers are given a chance to enter a one-way street the wrong way, more women than men will actually do so, but the survey by *Bild* simply counted wrongly entering cars and this is certainly no proof of their claim. For example, what if there were no men on the street at that time of the day? And in the case of the Swiss skiing resort, where almost all foreign tourists came from Germany, the attribution of abnormal dangerous behavior to this class of visitors is clearly wrong.

In terms of favorable events, *Der Spiegel*, on observing that, among foreigners, $P(\text{German tourist} \mid \text{skiing accident})$ was "large," concluded that the reverse conditional probability was also large, in particular, that being a German tourist increases the chances of being involved in a skiing accident:

$$P(\text{skiing accident} \mid \text{German tourist}) > P(\text{skiing accident}).$$

Similarly, *Hannoversche Allgemeine Zeitung* concluded from $P(\text{boy} \mid \text{bicycle accident}) = \text{large}$ that

$$P(\text{bicycle accident} \mid \text{boy}) > P(\text{bicycle accident})$$

and so on. In all these examples, the point of departure always was a large value of $P(A \mid B)$, which then led to the—possibly unwarranted—conclusion that $P(B \mid A) > P(B)$. From the symmetry

$$P(B \mid A) > P(B) \iff P(A \mid B) > P(A)$$

it is however clear that one cannot infer anything on A's favorableness for B from $P(A \mid B)$ alone, and that one needs information on $P(A)$ as well.

The British Home Office nevertheless once did so in its call for more attention to domestic violence (Cowdry, 1990). Among 1221 female murder victims between 1984 and 1988, 44% were killed by their husbands or lovers, 18% by other relatives, and another 18% by friends or acquaintances. Only 14% were killed by strangers. Does this prove that

$$P(\text{murder} \mid \text{encounter with husband}) > P(\text{murder} \mid \text{encounter with a stranger}),$$

that is, that marriage is favorable to murder? Evidently not. While it is perfectly fine to investigate the causes and mechanics of domestic violence, there is no evidence that the private home is a particularly dangerous environment (even though, as *The Times* mourns, "assaults ... often happen when families are together").

5. FAVORABLENESS AND SIMPSON'S PARADOX

Another avenue through which the attribute of favorableness can be incorrectly attached to certain events is Simpson's paradox (Blyth, 1973), which in our context asserts that it is possible that B is favorable to A when C holds, B is favorable to A when C does not hold, yet overall, B is unfavorable to A. Formally, one has

$$P(A \mid B \cap C) > P(A)$$

and

$$P(A \mid B \cap \bar{C}) > P(A),$$

yet

$$P(A \mid B) < P(A).$$

This paradox also extends to situations where $C_1 \cup \cdots \cup C_n = \Omega$, $C_i \cap C_j = \phi$ $(i \neq j)$. For real-life examples see Wagner (1982) or Krämer (2002, 2004), e.g.

One instance where Simpson's paradox (to be precise: the refusal to take account of Simpson's paradox) has been deliberately used to mislead the public is the debate on the causes of cancer in Germany. The official and fiercely defended credo of the Green movement has it that the increase in cancer deaths from well below 20% of all deaths after the war to almost 25% nowadays is mostly due to industrial pollution and chemical waste of all sorts. However, as Table 1 shows, among women, the probability of dying from cancer has actually *decreased* for young and old alike! Similar results hold for men.

These data refer only to mortality from cancer, not to the incidence of cancer, and therefore have to be interpreted with care. Still, the willful disregard of the most important explanatory variable "age" has turned the overall increase in cancer deaths into a potent propaganda tool.

If B is favorable to A, then by a simple calculation B is unfavorable to $\bar{A}$. However, B can still be favorable to subsets of $\bar{A}$. This is also known as Kaigh's (1989) paradox. In words: If knowing that B has occurred makes some other event A more probable, it makes the complementary event $\bar{A}$ less probable. However, we cannot infer that subsets of $\bar{A}$ have likewise become less probable.

Schucany (1989, Table 1) gives a hypothetical example where Kaigh's paradox is used to misrepresent the facts. Suppose a firm hires 158 out of 1000 applicants (among which 200 are black, 200 are Hispanic and 600 white). Of these, 38 non-whites and 120 whites are hired, amounting to 6.3% and 20% of the respective applicants. Being white is therefore favorable to being hired. But this does not imply that being in some non-white population is necessarily unfavorable to being hired. Assume for instance, that 36 of 38 non-whites who are hired are Hispanics. This implies that being Hispanic is likewise favorable for being hired. Although the selection rate for Hispanics is less than that for whites, we still have

$$\begin{aligned} P(\text{being hired} \mid \text{Hispanic}) &= 36/200 = 18\% \\ &> P(\text{being hired}) \\ &= 158/1000 = 15.8\%. \end{aligned}$$

Schucany (1989, page 94) notes: "Regardless of whether we call it a paradox, that such situations will be misconstrued by the statistically naive is a fairly safe bet."

A final and formally trivial example for faulty inferences from conditional probabilities concerns the inequality

$$P(A \mid B \cap D) > P(A \mid C \cap D).$$

Plainly, this does not imply

$$P(A \mid B) > P(A \mid C),$$

but this conclusion is still sometimes drawn by some authors. A German newspaper (quoted in Swoboda, 1971, page 215) once claimed that people get happier as they grow older. The paper's "proof" runs as follows: Among people who die at age 20–25, about 25% commit suicide. This percentage then decreases with advancing age; thus, for instance, among people who die aged over 70, only 2% commit suicide. Formally, one can put these observations as

$$\begin{aligned} &P(\text{suicide} \mid \text{age 20–25 } \underline{\text{and}} \text{ death}) \\ &\quad > P(\text{suicide} \mid \text{age} > 70 \ \underline{\text{and}} \text{ death}), \end{aligned}$$

and while this is true, it certainly does not imply

$$P(\text{suicide} \mid \text{age 20–25}) > P(\text{suicide} \mid \text{age} > 70).$$

TABLE 1

Probability of dying from cancer. Number of women (among 100,000 *in the respective age groups) who died from cancer in Germany*

Age	1970	2001
0–4	7	3
5–9	6	2
10–14	4	2
15–19	6	2
20–24	8	4
25–29	12	6
30–34	21	13
35–39	45	25
40–44	84	51
45–49	144	98
50–54	214	161
55–59	305	240
60–64	415	321
65–69	601	468
70–74	850	656
75–79	1183	924
80–84	1644	1587

Source: Statistisches Jahrbuch für die Bundesrepublik Deutschland.

In fact, a glance at any statistical almanac shows that quite the opposite is true.

Here is a more recent example from the U.S., where likewise $P(A \mid B)$ is confused with $P(A \mid B \cap D)$. This time the confusion is spread by Alan Dershowitz, a renowned Harvard Law professor who advised the O. J. Simpson defense team. The prosecution had argued that Simpson's history of spousal abuse reflected a motive to kill, advancing the premise that "a slap is a prelude to homicide" (see Gigerenzer, 2002, pages 142–145). Dershowitz, however, called this argument "a show of weakness" and said: "We knew that we could prove, if we had to, that an infinitesimal percentage—certainly fewer than 1 of 2,500—of men who slap or beat their domestic partners go on to murder them." Thus, he argued that the probability of the event K that a husband killed his wife if he battered her was small,

$$P(K \mid \text{battered}) = 1/2{,}500.$$

The relevant probability, however, is not this one, as Dershowitz would have us believe. Instead, the relevant probability is that of a man murdering his partner given that he battered her *and* that she was murdered,

$$P(K \mid \text{battered and murdered}).$$

This probability is about $8/9$ (Good, 1996). It must of course not be confused with the probability that O. J. Simpson is guilty; a jury must take into account much more evidence than battering. But it shows that battering is a fairly good predictor of guilt for murder, contrary to Dershowitz's assertions.

6. HOW TO MAKE THE SOURCES OF CONFUSION DISAPPEAR

Fallacies can sometimes be attributed to the unwarranted application of what we have elsewhere called "fast and frugal heuristics" (Gigerenzer, 2004). Heuristics are simple rules that exploit evolved mental capacities, as well as structures of environments. When applied in an environment for which they were designed, heuristics often work well, often outperforming more complicated optimizing models. Nevertheless, when applied in an unsuitable environment, they can easily mislead.

When a heuristic misleads, it is not always the heuristic that is to blame. More often than not, it is the structure of the environment that does not fit (Hoffrage et al., 2000). The examples we have seen here amount to what has elsewhere been called a shift of base or the base-rate fallacy (Borgida and Brekke, 1981). In fact, this environmental change underlies most of the misleading arguments with conditional probabilities.

Consider for instance the question: "What is the probability that a woman with a positive mammography result actually has breast cancer?" There are two ways to represent the relevant statistical information: in terms of conditional probabilities, or in terms of natural frequencies.

Conditional probabilities: The probability that a woman has breast cancer is 0.8%. If she has breast cancer the probability that a mammogram will show a positive result is 90%. If a woman does not have breast cancer the probability of a positive result is 7%. Take, for example, a woman who has a positive result. What is the probability that she actually has breast cancer?

Natural frequencies: Our data tells us that eight out of every 1000 women have breast cancer. Of these eight women with breast cancer seven will have a positive result on mammography. Of the 992 women who do not have breast cancer some 70 will still have a positive mammogram. Take, for example, a sample of women who have positive mammograms. How many of these women actually have breast cancer?

Apart from rounding, the information is the same in both of these summaries, but with natural frequencies the message comes through much more clearly. We see quickly that only seven of the 77 women who test positive actually have breast cancer, which is one in 11 (9%).

Natural frequencies correspond to the way humans have encountered statistical information during most of their history. They are called "natural" because, unlike conditional probabilities or relative frequencies, on each occurrence the numerical quantities in our summary refer to the same class of observations. For instance, the natural frequencies "seven women" (with a positive mammogram and cancer) and "70 women" (with a positive mammogram and no breast cancer) both refer to the same class of 1000 women. In contrast, the conditional probability 90% (the sensitivity) refers to the class of eight women with breast cancer, but the conditional probability 7% (the specificity) refers to a different class of 992 women without breast cancer. This switch of reference class easily confuses the minds of both doctors and patients.

To judge the extent of the confusion consider Figure 1, which shows the responses of 48 experienced doctors who were given the information cited above, except that in this case the statistics were a base rate

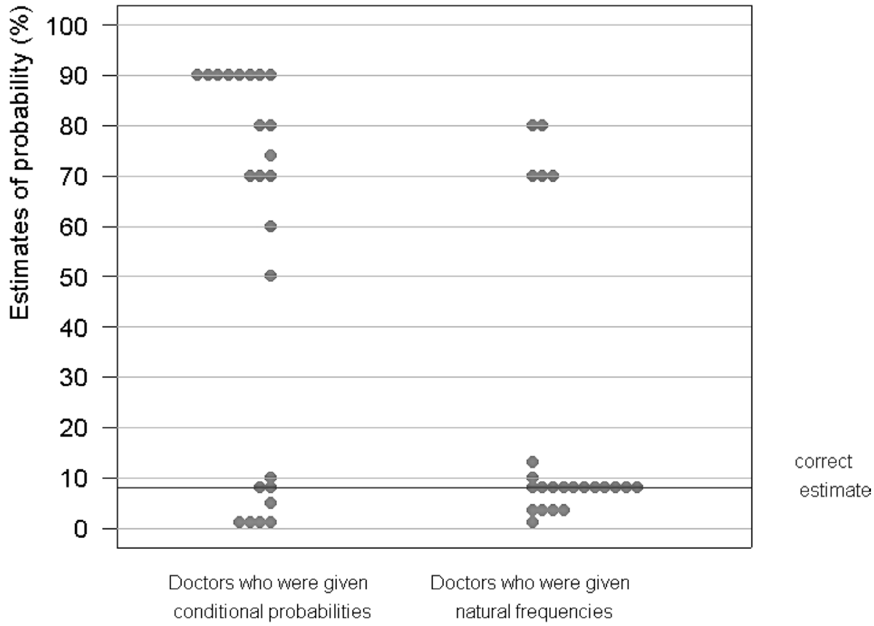

FIG. 1. *Doctors' estimates of the probability of breast cancer in women with a positive result on mammography* (*Gigerenzer,* 2002).

of cancer of 1%, a sensitivity of 80%, and a false positive rate of 10%. Half the doctors received the information in conditional probabilities and half received the data as expressed by natural frequencies. When asked to estimate the probability that a woman with a positive screening mammogram actually has breast cancer, doctors who received conditional probabilities gave answers that ranged from 1% to 90%; very few of them gave the correct answer of about 8%. In contrast, most of the doctors who were given natural frequencies gave the correct answer or were close to it. Simply converting the information into natural frequencies was enough to turn much of the doctor's innumeracy into insight. Presenting information in natural frequencies is therefore a simple and effective mind tool to reduce the confusion resulting from conditional probabilities.

ACKNOWLEDGMENTS

Research supported by Deutsche Forschungsgemeinschaft. We are grateful to Robert Ineichen for helping us to track down a reference to Leibniz, and to George Casella, Gregor Schöner, Michael Steele, Andy Tremayne and Rona Unrau for helpful criticism and comments.

REFERENCES

BAR-HILLEL, M. and FALK, R. (1982). Some teasers concerning conditional probabilities. *Cognition* **11** 109–122.

BECK-BORNHOLDT, H.-P. and DUBBEN, H.-H. (1996). Is the Pope an alien? *Nature* **381** 730.

BECK-BORNHOLDT, H.-P. and DUBBEN, H.-H. (1997). *Der Hund, der Eier legt—Erkennen von Fehlinformationen durch Querdenken.* Rowohlt, Hamburg.

BLYTH, C. R. (1973). Simpson's paradox and mutually favorable events. *J. Amer. Statist. Assoc.* **68** 746.

BORGIDA, E. and BREKKE, N. (1981). The base rate fallacy in attribution and prediction. In *New Directions in Attribution Research* (J. H. Harvey, W. J. Ickes and R. F. Kidd, eds.) **3** 63–95. Erlbaum, Hillsdale, NJ.

CHUNG, K.-L. (1942). On mutually favorable events. *Ann. Math. Statist.* **13** 338–349.

COWDRY, Q. (1990). Husbands or lovers kill half of women murder victims. *The Times*, April 14, p. 11.

D'ALEMBERT, J. ET AL. (1779). *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* **10**. J. L. Pellet, Geneva.

EDWARDS, A. W. F. (1996). Is the Pope an alien? *Nature* **382** 202.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

GIGERENZER, G. (2000). *Adaptive Thinking—Rationality in the Real World.* Oxford Univ. Press, New York.

GIGERENZER, G. (2002). *Calculated Risks*: *How to Know When Numbers Deceive You*. Simon and Schuster, New York. [British edition (2002). *Reckoning with Risk.* Penguin, London.]

GIGERENZER, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. In *Handbook of Judgement and Decision Making* (D. Koehler and N. Harvey, eds.) 62–88. Blackwell, Oxford.

GIGERENZER, G. and EDWARDS, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *British Medical J.* **327** 741–744.

GOOD, I. J. (1996). When batterer becomes murderer. *Nature* **381** 481.

GUILFORD, J. P. (1942). *Fundamental Statistics in Psychology and Education*. McGraw–Hill, New York.

HALLER, H. and KRAUS, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research* **7** 1–20.

HOFFRAGE, U., LINDSAY, S., HERTWIG, R. and GIGERENZER, G. (2000). Communicating statistical information. *Science* **290** 2261–2262.

KAIGH, W. D. (1989). A category representation paradox. *Amer. Statist.* **43** 92–97.

KRÄMER, W. (2002). *Denkste—Trugschlüsse aus der Welt des Zufalls und der Zahlen*, 3rd paperback ed. Piper, München.

KRÄMER, W. (2004). *So lügt man mit Statistik*, 5th paperback ed. Piper, München.

LESLIE, J. (1992). The doomsday argument. *Math. Intelligencer* **14** 48–51.

VON LEIBNIZ, G. W. (1887). *Die philosophischen Schriften*. (C. I. Gerhardt, ed.) **3**. Weidmann, Berlin.

MILLER, G. A. and BUCKHOUT, R. (1973). *Psychology*: *The Science of Mental Life*, 2nd ed. Harper and Row, New York.

NICKERSON, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods* **5** 241–301.

NUNNALLY, J. C. (1975). *Introduction to Statistics for Psychology and Education*. McGraw–Hill, New York.

SCHRAGE, G. (1993). Letter to the editor. *Math. Intelligencer* **15** 3–4.

SCHUCANY, W. R. (1989). Comment on “A category representation paradox,” by W. D. Kaigh. *Amer. Statist.* **43** 94–95.

SCHUCHARD-FISCHER, C., BACKHAUS, K., HUMMEL, H., LOHRBERG, W., PLINKE, W. and SCHREINER, W. (1982). *Multivariate Analysemethoden—Eine anwendungsorientierte Einführung*, 2nd ed. Springer, Berlin.

SWOBODA, H. (1971). *Knaurs Buch der modernen Statistik*. Droemer Knaur, München.

WAGNER, C. H. (1982). Simpson’s paradox in real life. *Amer. Statist.* **36** 46–48.

WYSS, W. (1991). *Marktforschung von A – Z*. Demoscope, Luzern.

ZIEGLER, H. (1974). Das Alibi des Schornsteinfegers—Unwahrscheinliche Wahrscheinlichkeitsrechnung in einem Mordprozeß. *Rheinischer Merkur* **39**.