Transparency in Risk Communication

Graphical and Analog Tools

ELKE KURZ-MILCKE,[a] GERD GIGERENZER,[b] AND LAURA MARTIGNON[a]

[a]*Institute of Mathematics and Computing, University of Education Ludwigsburg, Ludwigsburg, Germany*

[b]*Director, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany*

Why is it that the public can read and write but only a few understand statistical information? Why are elementary distinctions, such as that between absolute and relative risks, not better known? In the absence of statistical literacy, key democratic ideals, such as informed consent and shared decision making in health care, will remain science fiction. In this chapter, we deal with tools for transparency in risk communication. The focus is on graphical and analog representations of risk. Analog representations use a separate icon or sign for each individual in a population. Like numerical representations, some graphical forms are transparent, whereas others indiscernibly mislead the reader. We review cases of (1) tree diagrams for representing natural versus relative frequency, (2) decision trees for the representation of fast and frugal decision making, (3) bar graphs for representing absolute versus relative risk, (4) population diagrams for the analog representation of risk, and (5) a format of representation that employs colored tinker cubes for the encoding of information about individuals in a population. Graphs have long enjoyed the status of being "worth a thousand words" and hence of being more readily accessible to human understanding than long-winded symbolic representations. This is both true and false. Graphical tools can be just as well employed for transparent and nontransparent risk communications.

Key words: **risk communication; risk perception; graphics; natural frequencies; fast and frugal heuristics**

Women have gone through many "pill scares" since the introduction of the contraceptive pill in the 1960s. One of these concerned the third-generation pill that contained desogestrel and gestodene. In the mid-1990s, the British press reported the results of a study that women who took this contraceptive pill increased their risk of thromboembolism by 100%. *Thromboembolism* means blockage of a blood vessel by a clot and can lead to fatal strokes. Hearing the bad news, thousands of British women panicked and stopped taking the pill, which led to a wave of unwanted pregnancies. But what did the study in fact show? Out of every 14,000 women who did not take the pill, one had thromboembolism, and out of every 14,000 who took it, this number increased from one to two. That is, the absolute risk increase is 1 in 14,000, but the relative increase is 100%.[1] Most of these women, like the majority of people, had never learned the difference between absolute and relative risks and thus could be easily frightened by the 100% figure. Estimates are that more than 10,000 British women had abortions as a consequence of the press release.

The pill scare illustrates the public's lack of understanding of statistical information as well as the emotional and physical consequences. Who would be interested in the public not understanding risks? Typically, the pharmaceutical industry is blamed. In the present case, however, both women and the pharmaceutical industry were hurt; the only persons who profited were the journalists who created front-page headlines. If the journalists had reported the absolute risk, their editors would not have been impressed and women would have been spared an unnecessary fright. Yet, there are other potential culprits. Scientists who publish similar studies often report the benefits or risks of treatments in relative risks since they suggest bigger effects and are more likely to be published. This policy has been and continues to be tolerated by medical journals, which seem to be conflicted between reporting results in a transparent way or as relative risks, as these larger percentages get the journal into the news.

Address for correspondence: Elke Kurz-Milcke, Institute of Mathematics and Computing, University of Education Ludwigsburg, Reuteallee 46, 71634 Ludwigsburg, Germany. Voice: +49-71-411-40-729.
kurzmilcke@ph-ludwigsburg.de

Ann. N.Y. Acad. Sci. 1128: 18–28 (2008).
doi: 10.1196/annals.1399.004

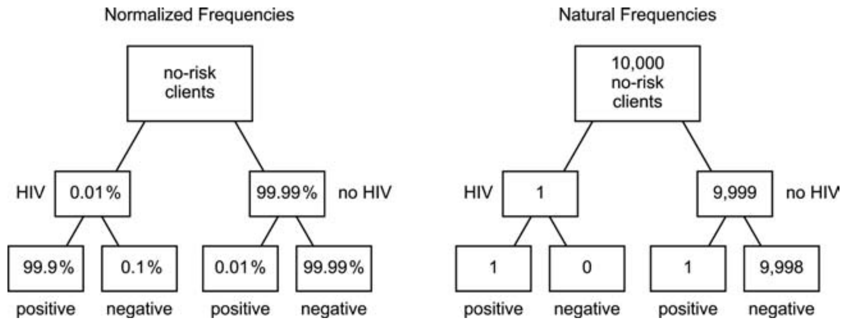

FIGURE 1. Two kinds of frequency trees for HIV testing: relative frequencies (left), which are nontransparent for many people, and natural frequencies (right), which are transparent.

Why is it that the public can read and write but only few understand statistical information? Why are elementary distinctions, such as that between absolute and relative risks, not better known? One answer is historical. Statistical information was once a state secret, as reflected in the origins of its name. Not until 1830 did statistical information become accessible to the public on a large scale, hand-in-hand with the democratization of various countries. Open access to information is one pillar of democracy, but it is not sufficient as information also needs to be transmitted transparently. The British women received the correct information but in a form that was confusing. They may have thought they understood what 100% meant, but they were misled.

The pill scare is far from being a one-off event. A glance into medical journals, information brochures for cancer screening, and doctor–patient interaction reveals that transparent communication is the exception rather than the norm. At a time when people are talking about informed consent, shared decision making, and evidence-based medicine, transparency has become essential. Without transparent risk communication, informed consent is impossible.

In this chapter, we deal with tools for transparency in risk communication. The focus is on graphical and analog representations of risk, but we will see that the old saying "a picture is worth a thousand words" does not always hold true. Like numerical representations, some graphical forms are transparent whereas others indiscernibly mislead the reader. We do not deal with the conflicts of interests that cause and tolerate the use of nontransparent tools.[2] Yet, understanding and implementing transparency will affect the institutions and individuals who promote nontransparent risk communication. If patients, physicians, and the general public understand how to translate nontransparent into transparent communication, it will become increasingly difficult to manipulate public opinion. In this sense, not only institutions but also representations can carry authority.[3]

Trees: Natural versus Relative Frequency

From Darwin's tree of life to genealogical family trees, tree diagrams have been used for representing relationships since time immemorial.[4] Tree diagrams also play a prominent role in probability and statistics. They are intuitive in that their structure suggests how to read them. According to our conventional way of reading, we move from the top to the bottom of a page. As a consequence, tree diagrams are usually inverted relative to what their metaphorical name suggests; that is, their branches extend downwards.

Not every tree is a transparent representation, however. Consider the case of a newly married and pregnant woman in California whose doctor advised her to take a routine HIV test. A week later, her doctor phoned to say that she had tested positive in the ELISA and the Western blot tests, which meant that she was definitely infected with HIV. He also told her to inform her husband and family immediately. That was the worst day of her life. The next morning, she began to deliberate and recalled a study about the test's reliability. FIGURE 1 (left) shows the way the relevant information is typically communicated, here in the form of a tree. Only 0.01% of people who are not engaged in risky behavior, such as intravenous drug use, are infected. Of those, 99.9% test positive; this percentage is called the sensitivity of the test. At the same time, 0.1% wrongly test negative. Among those who are not infected, 99.99% correctly test negative, a percentage that is called the specificity of the test. Only 0.01% wrongly test positive; this is the false positive rate. Given this information, what is the probability that a person who tested positive is actually HIV-infected?

	HIV	no HIV	
test positive	1	1	2
test negative	0	9,999	9,998
	1	9,999	

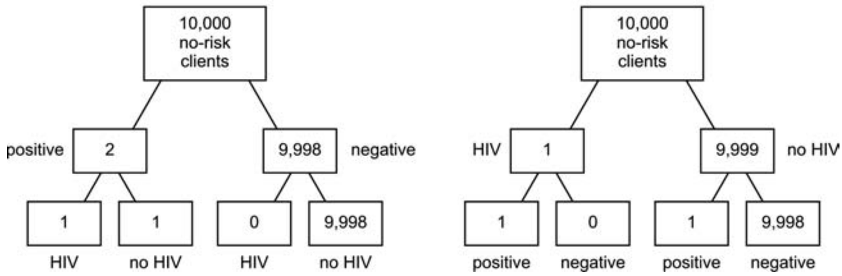

FIGURE 2. A contingency table for HIV testing and the two resulting natural frequency trees.

Most of our readers will be confused by this tree, as were a majority of professional AIDS counselors.[5] The typical answer is that the probability is 99.9% or 99.99% that the woman actually is infected. Such a tree is called a relative frequency tree, and it baffles people even though it represents accurate information.

FIGURE 1 (right) shows a natural frequency tree, which represents the same information in a transparent way. Out of every 10,000 people, one is likely to be infected, and this person is virtually certain to test positive. Of those who are not infected, we expect that one will nevertheless test positive. With this transparent tree, we can see the answer to the above question immediately: There are two people who test positive, and only one of them is infected. Thus, the chances that the woman who tested positive truly was infected are about 50:50.

In this particular case, the Californian woman, together with her husband, went to another testing center the next day, and both have tested negative ever since.

Why do trees with natural frequencies foster insight? The reason is that the natural frequency representation does part of the computations. If one wants to compute the probability $p(\text{HIV}\,|\,\text{positive})$ from the relative frequency tree, one would have to perform the following mental calculations:

$$p\,(\text{HIV}|\text{positive}) = p\,(\text{HIV})p\,(\text{positive}|\text{HIV})/[(p\,(\text{HIV})$$
$$p\,(\text{positive}|\text{HIV}) + p\,(\text{no HIV})p\,(\text{positive}|\text{no HIV})]$$

This equation states that the probability someone has the HIV virus, given that they test positive, is equal to the probability that they have the virus multiplied by the probability that they test positive, given they have the virus, divided by the sum of that number and the probability they do not have the virus multiplied by the probability they test positively, given they do not have the virus. This results in

$$0.01\% \times 99.9\%/(0.01\% \times 99.9\%$$
$$+\,99.99\% \times 0.01\%) \approx 50\%.$$

In contrast, the natural frequency tree reduces these computations to:

$$p\,(\text{HIV}|\text{positive}) = 1/(1+1)$$

The facilitating effect of natural frequency representations also becomes apparent when they are compared with standard contingency tables. The frequency of co-occurrence of two binary features or categories in a population is represented by 2×2 contingency tables. Alternatively, co-occurrence may be represented by natural frequency trees. Note that each natural frequency tree represents either the rows or the columns of the corresponding table. Thus, to fully represent a 2×2 contingency table, two natural frequency trees are required.

FIGURE 2 shows a contingency table for the case of HIV testing with the corresponding natural frequency trees. The columns show the figures for health status, the rows the figures for test outcome. On their first level, the corresponding natural frequency trees partition the population of 10,000 individuals either by test result (FIG. 2, left) or by health status (FIG. 2, right). Thus, the numbers on the first level correspond to the marginal frequencies of either the rows or columns in the table. On their second level, the trees show the exact same partitioning of the total number of individuals.

These numbers naturally correspond to the cells in the contingency table as well. When considering a test's sensitivity, that is, the number of those infected who correctly test positive, and its specificity, that is, the number of individuals not infected who correctly test negative, the initial criterion for partitioning is health status (as in FIG. 1). By contrast, when interpreting test results, the initial criterion is the test outcome. Trees that partition the same population by different criteria can also be combined in double-tree structures.[6]

As noted above, what is remarkable about HIV testing of persons with no known risk of infection is that only one of two individuals who test positive is actually infected with the virus. Tree diagrams can assist in understanding how such a situation can arise even when a test demonstrates remarkably good sensitivity and specificity. In this manner, natural frequency trees can support Bayesian inversion; that is, inferring the probability of health status given test outcomes from the probability of test outcomes given health status. Natural frequency trees explicate this inference by switching between criteria for the partitioning of a population, in this case the 10,000 persons with no known risk of HIV infection.

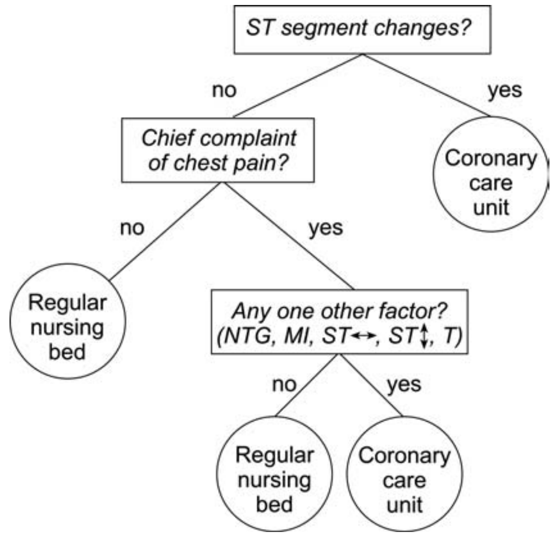

FIGURE 3. A fast and frugal tree for coronary care unit allocation (based on Ref. 7). ST segment changes are a particular anomaly of the electrocardiogram. The factor NTG signifies a history of nitroglycerin use for chest pain, the factor MI myocardial infarction, ST ↔ and ST ↕ signify ST segment barring, and T stands for T waves with peaking or inversion.

Trees: Fast and Frugal Trees versus Full Trees

Natural frequency trees can be extended without much difficulty to situations in which the co-occurrence of more than two binary variables is considered, for example, multiple tests for a condition.[7] Thus, the question arises whether all of the branches in more extensive trees are of relevance or whether it is even practical to consider all possible combinations of tests or cues in a situation. By way of illustration, consider the decision tree for treatment allocation designed by Green and Mehr[8] in FIGURE 3. The diagnostic problem for which the tree is designed is the following: A person is rushed into a hospital with severe chest pains. The doctors have to decide whether the patient should be sent into the coronary care unit or into a regular bed with telemetry. This can be a life-or-death decision since, if the patient has a heart attack, he or she should be in the care unit, otherwise in a regular bed. A number of well-established criteria are relevant to the decision; nevertheless, the question remains how to implement a decision procedure that is sufficiently fast and makes economical use of the potential informational sources. Another way of phrasing this challenge is how to determine the most adequate ordering of a selection of cues, among them symptoms, test results, and indicators from the patient's medical history. This question is relevant with trees that are "pruned" and whose branches do not all follow through to the final level.

A fast and frugal tree is such a pruned tree. Fast and frugal trees have an exit on every level, resulting in $n + 1$ exits for binary trees, where n is the number of cues or indicators considered. The particular problem that Green and Mehr faced was that doctors in a Michigan hospital sent 90% of all patients with severe chest pains into the coronary care unit, which led to overcrowding, decreasing the quality of care and increasing costs. In addition, it provided a health risk for the patients since the care unit is one of the most dangerous places in the hospital (patients can die from viruses circulating there).

The fast and frugal tree consists of three questions or cues, one of them a compound cue based on a disjunction of four well-established diagnostic indicators.[9,10] Green and Mehr's selection and ordering of cues was very successful: Their data set consisted of 89 admissions; none of the patients sent to a regular bed exhibited infarction. In other words, Green and Mehr's decision procedure did not result in any false negatives. The procedure did, however, send 35 patients not exhibiting infarction to the coronary care unit, corresponding to a false positive rate of 35 out of 74. Although this decision procedure showed the desirable performance of minimizing false negatives, resulted in a much lower number of admissions to the care unit, and was structured in the fashion of a fast and frugal binary decision tree, the question of how the implemented cue ordering had been determined by the authors remained essentially unspecified. It can be assumed that the devised order was at least informed by the clinicians' knowledge of a logistic regression-based instrument available for this type of clinical decision making.[10]

Based on the same data and cue selection, Martignon *et al.*[9] worked out a more formalized procedure for constructing fast and frugal decision trees for the diagnostic problem of admission to the coronary care unit. Their approach was based on the values of sensitivity, specificity, positive validity, and negative validity for the individual cues. By combining cues according to principles that took account of these performance values, they arrived at cue orderings for several fast and frugal decision trees. Notable in this context was that more than one of these trees performed comparably to the fast and frugal procedure proposed by Green and Mehr. The following is an example of an approach that resulted in a different but in terms of its performance equivalent ordering of cues to that of Green and Mehr and that was exclusively based on the cues' positive and negative validities[9]:

Cue ordering approach: maximize (positive validity; negative validity)

- Begin with the cue that shows maximal positive validity.
- All remaining cues are ranked by the maximum of positive validity/negative validity.

Decision rule for the corresponding fast and frugal binary decision tree:

- At each level the decision is made so that the branch with the larger number of individuals is the exit on that level; a positive cue value on the exit branch sends the patients to the coronary care unit, a negative value to a regular bed.
- The next cue in the cue ordering is applied to the branch with the smaller number of individuals.
- Continue until all of the rank-ordered cues are applied.

The principles for establishing this and other cue orderings are simple in that they do not take account of conditional dependencies among cues. In other words, for establishing the respective performance values of the individual cues, these cues are treated as a singular source of information about the respective criterion, for instance, occurrence of infarction. Fast and frugal decision trees can show desirable levels of performance while exhibiting an especially economical (i.e., frugal) use of informational sources by limiting decision procedures to a comparatively small number of cues and establishing orderings among cues that are based on performance values for individual cues.[11]

An important question is whether fast and frugal decision trees can help transparency in risk communication. Clearly, all decision trees specify a decision procedure that can be visually communicated. Thus, the respective tree diagram can help with describing, understanding, and learning a decision procedure. Fast and frugal trees have the advantage that they generally employ a limited number of levels (compared to, for instance, full Bayesian trees), which makes it easier to understand, memorize, and communicate them. However, it is also important to make transparent how these trees are constructed. This aspect is especially relevant when introducing a decision procedure to an expert community that may be very reluctant to work with a "black-boxed" decision device. Transparency cannot happen without also specifying normative issues about design criteria, such as what are considered good arguments for excluding cues from consideration in a decision procedure (for studies of ecological rationality, see Ref. 11).

Bars: Absolute versus Relative Risk Bars

Technically, a bar graph represents the frequency of the events in question by the height of the bars in the graph. A histogram, by comparison, represents frequency by area. Bar graphs strongly afford a fairly automatic and precise kind of perceptual comparison of heights that comes with a high degree of subjective confidence. We humans have a strong perceptual capacity for the comparison of lengths in two-dimensional displays within our visual field, which we can quickly scan at a glance. Nevertheless, this capacity can be exploited; for example, bars can be constructed in a way that differences appear much larger than they are.

FIGURE 4 shows two bar graphs that represent the same information on risks of stroke and major bleeding in patients with atrial fibrillation but in two different ways (adapted from Ref. 12, p. 743). In particular, the graph on the left-hand side suggests comparisons of incidences with and without treatment. Specifically, it turns out that the bars for the categories "no treatment," "Aspirin," and "Warfarin" show perceptually comfortable, that is, clearly differentiable, differences in height. Everything seems to be present, the two

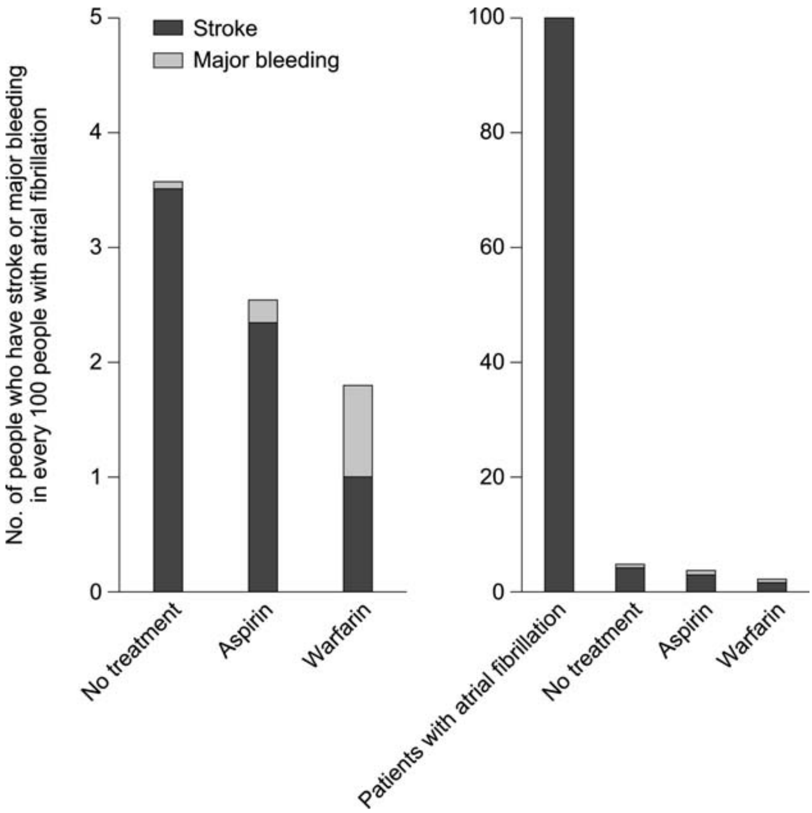

FIGURE 4. Two bar graphs representing the same benefits of treatment in two different ways. The absolute effect of Aspirin and Warfarin becomes transparent in the bar graph on the right when the reference population is included (based on Ref. 12).

treatments and the control group a perfect representation, so it may seem. However, this representation invites the same confusion as when reporting relative risk reduction (in the contraceptive pill scare). In numbers, the risk that a patient with atrial fibrillation will experience stroke is reduced by approximately 70% when taking Warfarin and by half of that when taking Aspirin. The bar graph on the right-hand side, in contrast, gives a visualization of absolute risk reduction. It shows a bar indicating a reference group of 100 people with atrial fibrillation. In terms of absolute risk reduction, the corresponding figures for stroke are: Three to four out of every 100 people with atrial fibrillation who do not get treatment are expected to experience stroke. By comparison, one out of every 100 people with atrial fibrillation who take Warfarin is also expected to experience stroke, and slightly more than two out of 100 who take Aspirin. With this additional information, the effect of the treatments becomes transparent.

Many people believe that pictorial representations are more informative and easier to understand than numbers. That is an error. Graphs can be equally misleading, for instance, by leaving reference classes implicit and by obscuring relevant comparisons.

Analogs: Population Diagrams

Population diagrams represent frequency in analog fashion, e.g., by using a population of analogous icons, each representing an individual, rather than by using number symbols or bar height. By this one-to-one match between individual and icon, population diagrams invite identification. To a greater extent than with trees and bars, the reader can imagine being one of the individuals in the diagram.

Consider the population diagrams in FIGURE 5 out of an editorial in *The New England Journal of Medicine* in which Elmore and Gigerenzer[13] addressed the issue of transparent risk communication in communicating study-based evidence on the risks of breast cancer. The question was whether various findings on breast biopsies predict the occurrence of breast cancer after 15 years. The following is a summary statement of the results described in the original study:[13]

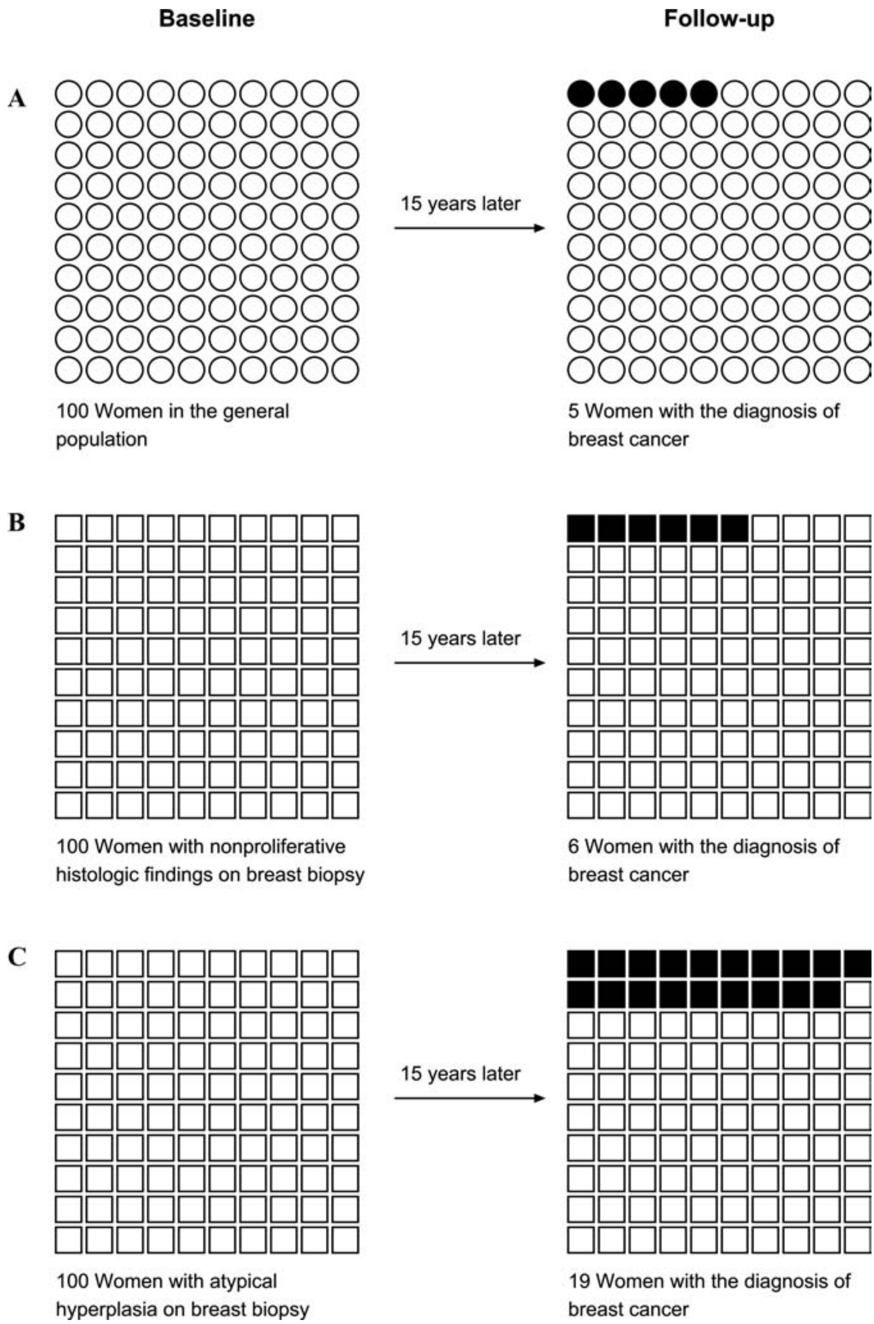

FIGURE 5. A population diagram representing the number of women who will be diagnosed with breast cancer in the next 15 years. Each circle represents a woman in the general population, and each square represents a woman who has undergone a biopsy. The black circles and squares represent women with a diagnosis of breast cancer after an average of 15 years of follow-up (based on Ref. 13).

As compared with women in the general population, women with nonproliferative findings on breast biopsy had a relative risk of breast cancer of 1.27 and those with atypia a relative risk of 4.24.

The results are framed in terms of relative risk increase and are nontransparent for most patients and physicians. FIGURE 5 presents the same results transparently by mapping individual patients onto individual icons. Let us assume the perspective of a woman who underwent biopsy and learned about a nonproliferative histological finding. There are obviously three major places of action in this figure, namely the three population diagrams on the right. In each, a relatively small number (on closer inspection $n < 20$) of icons is set off from the rest, while each of the diagrams (on closer inspection a 10×10 array) includes the same number of icons. The arrows indicate a before-and-after evaluation. The filled-in icons represent women with breast cancer and thus a development from not having cancer to having it. The three panels indicate three situations or groups of women. They help to localize our protagonist in the center panel. It is clear that one group incurs a higher risk of having cancer than the one she now finds herself in, but there is also a group of women who incur less of a risk. On close inspection, the middle panel has six filled-in icons, the upper panel five. The upper panel represents women in the general population as a base line for comparison (note that they are represented by circles rather than the squares representing the two clinical populations). Our protagonist can see that the prognosis for women like her of actually developing cancer over the next 15 years is remarkably similar to the prognosis for women in the general public, whereas the relative risk increase of 1.27 might not have conveyed this information in clear terms. Although she is no doubt relieved to be not one of those in the third panel, she may, nevertheless, note that most of these women will not develop breast cancer over the next 15 years. Letting this picture sink in and understanding the amount of risk increase will have an impact on our protagonist's anxiety and other emotional reactions.

In FIGURE 5, individuals are represented by circles and squares. One step beyond this is to use icons that resemble the objects represented; for instance, when people are the objects to shape the icons using salient properties, such as identifiable clothing for a professional group or a particular body shape for small children. The Viennese philosopher and socialist Otto Neurath[14] thought that these types of icons, which he called Isotypes, could become a visual language that allows the general public to understand the statistical structure of the world. Consider the information sheet in FIGURE 6 about the Triple Test, which is used to determine increased risks for Down syndrome and for a neural tube defect during pregnancy.[15] In this figure, individuals are not represented by squares or circles, but the icons are instead modeled after the pregnant female body.

The diagram presents a "crowd" of 1000 pregnant women who participated in screening in a 50×20 array. Most individuals are shown in black and make up the happy part of this crowd for they have been identified as not showing increased risks. Two groups stand out from this crowd. For the 90 individuals 40 individuals in the upper left-hand corner (including the two unshaded individuals associated with that group), the test indicates an increased risk of Down syndrome; for the in the upper right-hand corner (again, including two unshaded individuals), an increased risk of a neural tube defect. To persons participating in medical screening, standing out from the "screening crowd" generally signals a warning. Yet, only the children of those not shaded actually have Down syndrome and neural tube defect, respectively. All of the other individuals in these two groups are false positives. We can see how many false alarms this test creates. Among the black individuals, the unshaded icon indicates a miss. This woman received a negative test result, but her child shows one of the two clinical conditions. Thus, a pregnant woman is able to see from this diagram that a positive result on screening with the Triple Test does not imply Down syndrome or a neural tube defect, or even their likelihood. Population diagrams allow individuals to "see" what their chances are. This type of seeing is backed by human understanding of group membership and intuitions about what it means to stand out from a crowd.[16]

Analogs: Tinker Cubes

Tinker cubes are a medium for representation that is also used in the mathematics classroom.[17,18] As with population diagrams, tinker cubes can represent frequency in analog fashion cubes represent individuals. Those shown in FIGURE 7 have a side length of about 1 cm and are made of variably colored plastic. Five sides of these cubes have a small hole punched in their middle, and the sixth side has a small stud that can be connected to another cube. When presented with this kind of material, people tend to spontaneously assemble and sort cubes.

The charm of composites, such as those in FIGURE 7, is that they can code the co-occurrence of features characterizing individuals in a population. In this case, we speak of conjunctions. Consider the container in FIGURE 8, which holds 1000 conjunctions of the kind shown in FIGURE 7.

This population of conjunctions represents 1000 women who participated in mammography screening. Each individual in this "screening urn" is represented by a tinker cube indicating health status and a tinker cube indicating test result. Specifically, from left to right in FIGURE 7: (1) a red and blue conjunction represents a woman without cancer and a negative result

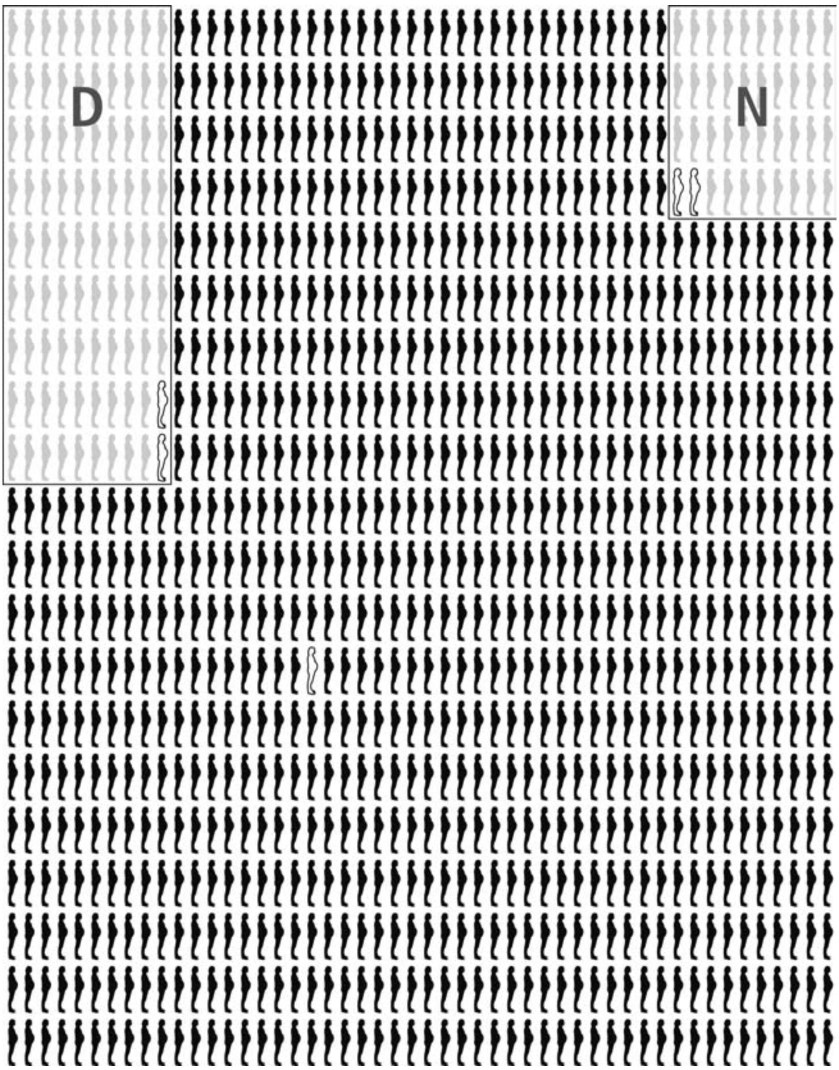

FIGURE 6. A population diagram representing 1000 pregnant women who undergo the Triple Test (after a diagram by W. Holzgreve, R. Schloo, & P. Miny, patient information leaflet; see for instance Ref. 15). The 90 individuals forming the group in the upper left-hand corner test positive on Down syndrome (*D*), the 40 individuals in the upper right-hand corner test positive on a neural tube defect (*N*). Yet, most positive tests are false positives; only the two individuals shown in white among each group have the respective clinical condition. Among those who test negative, i.e., the individuals shown in black, there is one pregnant woman whose child will nonetheless have one of these clinical conditions.

on mammography screening, the most desirable status; (2) a red and white conjunction represents a woman without cancer and a positive test result, a false alarm; (3) a green and white conjunction represents a women with breast cancer and a positive test result, a hit; (4) a green and white conjunction represents a woman with breast cancer and a negative test result, a miss. What is the distribution of these four types of individuals in the screening urn? There are 922 red and blue conjunctions, 70 red and white conjunctions, seven green and white conjunctions, and a single green and blue conjunction (numbers as in Ref. 2, p. 45). Imagine a woman who participates in breast cancer screening and whose mammography comes back positive: What is the probability of actually having breast cancer given a positive test in this situation? In order to answer this question, we only need to consider those conjunctions from the screening urn that carry a white

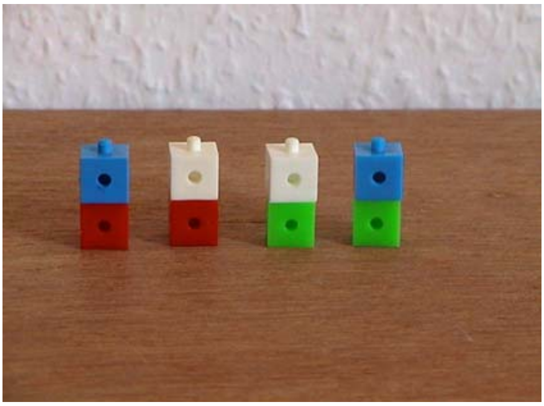

FIGURE 7. Conjunctions consisting of two tinker cubes each. The color of the cubes is used for the coding of two binary variables, namely *health status* and *test result on mammography screening*. Each conjunction represents an individual, in this case a woman participating in mammography screening. Red signifies *no cancer*, green signifies *cancer;* blue signifies a *negative test result*, white a *positive test result*. Note that this color assignment was chosen because of the differing numbers of tinker cubes of each color available in the collection of didactical materials at the University of Education in Ludwigsburg, Germany. The composition of the corresponding *mammography screening urn* shown in FIGURE 8 demanded vastly larger numbers of tinker cubes in the colors that represented *no cancer* for *health status* and *negative* for *test result* than for the complementary values of these variables (see text for the corresponding numbers). (In color in *Annals* online.)

FIGURE 8. *Mammography screening urn* consisting of 1000 conjunctions where each conjunction represents a woman participating in mammography screening according to her *health status* and her *test result*. (see text and caption of Fig. 7 for color assignments). (In color in *Annals* online.)

tinker cube. White tinker cubes represent women with a positive result on mammography screening. Sorting the conjunctions in the screening urn for those carrying a white cube leads to a total of 77 conjunctions, 70 red and white conjunctions and seven green and white conjunctions. Only those carrying a green cube have breast cancer. In other words, only seven out of 77 individuals with a positive mammography actually have cancer. What, then, is the probability of actually having breast cancer given a positive test? We can imagine blindly drawing conjunctions many times from an urn containing only those that include a white tinker cube, that is, the 77 specified conjunctions. Each time a single conjunction is drawn, the color of the second cube is noted, and the pair of cubes is returned to the urn. We can expect to draw green and white conjunctions from this urn with a probability of 7/77.

Tinker cube representations invite simulative reasoning. Even children 9 years of age and younger can imagine drawing pairs of cubes from an urn many times, and they can reason in adequate ways about series of drawings when given representations that help them record the outcomes.[19] Also, the ease with which conjunctions are sorted, added, and removed from a population of tinker cubes allows for the ready manipulation of populations. These manipulations need not be actually carried out or observed but may be simply imagined or observed in arrangements or distributions. For instance, sorting by just one color (e.g., the white cubes in the described screening urn) prepares one for understanding a conditional frequency: Given a white cube, what is the frequency of red or green, respectively? These relative frequencies can be counted or an impressionistic judgment about them can be formed when looking at the respective sorting. Clearly, we can sort by any of the four colors of tinker cubes or by the four types of conjunctions (see above for figures referring to the sorting by conjunctions). As a final example, let us assume that the green tinker cubes are isolated. This leads to a relatively small number of conjunctions, where only eight of the 1000 individuals have cancer. Seven of these are paired with a white tinker cube and one with a blue cube. This gives us the test's sensitivity as the relative frequency that women with breast cancer (green) have a positive result on mammography screening (white), namely in seven out of eight cases. More often than not, the test is accurate in this way. Tinker cube representations can help to differentiate between conditional frequencies referring to different reference classes. In this case, we focused on the individuals who have cancer; at other times, as described above, we may be interested in those who test positive. The difficulty with symbolic representations, namely confusing the conditional probabilities of p(positive | cancer) with those of p(cancer | positive), is a nonissue with the tinker cubes.

There is No Informed Society without Transparency

The graphical representations discussed in this chapter exemplify how transparency can be achieved and also how it can be prevented if medical societies and doctors continue to use representations that the human mind cannot easily digest. Graphs have long enjoyed the status of being "worth a thousand words" and hence of being more readily accessible to human understanding than long-winded symbolic representations. This is both true and false. Graphical tools can be just as well employed for transparent and nontransparent risk communications. Understanding this is one important step toward becoming an informed society.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Jain, B.P., McQuacy, H. & Moore, A. 1998. Number needed to treat and relative risk reduction. Ann. Intern. Med. **128:** 72–73.
2. Gigerenzer, G. 2002. Calculated risks: How to Know When Numbers Deceive You. Simon & Schuster. New York, (UK version: Reckoning with Risk, Penguin).
3. Kurz-Milcke, E. 2004. The authority of representations. *In* Experts in Science and Society. E. Kurz-Milcke & G. Gigerenzer, Eds.: 281–301. Kluwer/Plenum. New York.
4. Bredekamp, H. 2005. Darwins Korallen: Frühe Evolutionsmodelle und die Tradition der Naturgeschichte [Darwin's corals: Early models of evolution and the tradition of natural science]. Klaus Wagenbach. Berlin.
5. Gigerenzer, G., U. Hoffrage & A. Ebert. 1998. AIDS counselling for low-risk clients. AIDS Care **10:** 197–211.
6. Wassner, C., S. Krauss & L. Martignon. 2002. Muss der Satz von Bayes schwer versätndlich sein? [Is Bayes's equation necessarily difficult to understand?] Praxis der Mathematik **44:** 12–16.
7. Krauss, S., L. Martignon & U. Hoffrage. 1999. Simplifying Bayesian inference: the general case. *In* Model-based Reasoning in Scientific Discovery. L. Magnani, N.J. Nersessian & P. Thagard, Eds.: 165–180. Kluwer Academic. New York.
8. Green, L.A., & D.R. Mehr. 1997. What alters physicians' decisions to admit to the coronary care unit? J. Fam. Practice **45:** 219–226.
9. Martignon, L., O. Vitouch, M. Takezawa & M.R. Forster. 2003. Naïve and yet enlightened: from natural frequency trees to fast and frugal decision trees. *In* Thinking: Psychological Perspectives on Reasoning, Judgement and Decision Making. D. Hardmann & L. Macchi, Eds.: 189–211. John Wiley & Sons. Chichester, UK.
10. Gigerenzer, G. 2007. Gut Feelings: The Intelligence of the Unconscious. Viking. New York. (UK: Allen Lane/Penguin).
11. Gigerenzer, G., P.M. Todd & the ABC Research Group. 1999. Simple Heuristics that Make Us Smart. Oxford University Press. New York.
12. Gigerenzer, G. & A. Edwards. 2003. Simple tools for understanding risks: from innumeracy to insight. BMJ **327:** 741–744.
13. Elmore, J.G. & G. Gigerenzer. 2005. Benign breast disease—The risks of communicating risk. N. Engl. J. Med. **353:** 297–299.
14. Neurath, O. 1945. Health education through Isotype. Lancet (Aug. 25, 1945): 236–237.
15. Villars, P., Pränatal-Tests zur Risikoabschätzung von kindlichen Erkrankungen oder Erbgutstörungen, http://www.villarsgyn.ch/praenataltest.htm, 04/30/2006, date accessed: 05/25/2007.
16. Wang, X.T. 1996. Domain-specific rationality in human choices: violations of utility axioms and social contexts. Cognition **60:** 31–63.
17. Kurz-Milcke, E., & L. Martignon. 2006. Lebendige Urnen und ereignisreiche Bäume: Überlegungen und Versuche zu einer Didaktik der Stochastik in der Grundschule. [Lively urns and eventful trees: Developing a didactics of stochastics in elementary school] Anregungen zum Stochastikunterricht **3:** 182–203.
18. Martignon, L., K. Laskey & E. Kurz-Milcke. 2007. Transparent Urns and Colored Tinker-cubes for Natural Stochastics in Primary School. Proceedings of CERME 5-2007, Conference of the European Society of Reasearch in Mathematics Education- Group V -3.
19. Wollring, B. 1994. Qualitative empirische Untersuchungen zum Wahrscheinlichkeitsverständnis bei Vor- und Grundschulkindern [Qualitative empirical studies of the concept of probability in preschool and elementary school children]. Habilitation thesis, Westfälische Wilhelms-Universität Münster, Germany.