Probability, problems, and paradoxes pictured by eikosograms

R.W. Oldford
University of Waterloo

April 21, 2003

Abstract

Eikosograms are diagrams which embed the rules of probability and can be used to understand and to explore the probabilistic structure involving one or more categorical variables. Rectangular areas correspond to probabilities and can be used to calculate their numerical value and to determine the Bayes relation.

Eikosograms are used here to resolve a series of well known problems and paradoxes including a novel version of the secretary problem (here a gas station problem), a preference paradox (the Cherry pie paradox), a twice reversing Simpson's paradox, the two-envelope paradox, the Monty Hall problem, and the prisoner's dilemma (or three prisoner problem). Two new variations on the last two problems are also introduced via eikosograms.

Additionally, situations where probability must to be understood for informed decision making are shown to be considerably simplified by visually examining the relevant eikosograms. The contextual situations for decision making used here are: choosing between medical treatments in light of the probabilities of various outcomes, understanding characteristics of medical (or other) tests including sensitivity, specificity or opacity, and their relationship to the probability of false positive or false negative testing (in the context of osteoporosis here), and the proper understanding of the reliability of eye witness testimony (the famous green-cab, blue-cab hit and run). These situations are described in full and exercise a fundamental understanding of probability – marginal, conditional and the Bayes relation – via simple visual display.

The situations, problems and paradoxes have been selected so that working through them in sequence brings about a relatively deep understanding of probability in a simple visual way. Throughout, symbolic and diagrammatic reasoning operate in parallel by way of the eikosograms and the problems and paradoxes are resolved with little or no mathematics.

Keywords: base rates, Bayes rule, cherry pie paradox, eyewitness testimony, gas station problem, graphical display, judgement under uncertainty, medical testing, Monty Hall, opacity, outcome trees, ORAI, osteoporosis, prevalence, prisoner's dilemma, probability displays, secretary problem, sensitivity, Simpson's paradox, specificity, teaching probability, two envelope paradox,

1 Introduction

The eikosogram is a simple graphical display showing probability as the area of rectangles which tile a unit square. Cherry and Oldford (2003) show that the diagram embeds axioms of probability based on random variables and argue that the eikosogram is coincident with the meaning of probability, independent of its application. The argument is also made that eikosograms should replace the Venn diagram as the diagram of choice for understanding and defining probability.

In this paper, I show how eikosograms are able to display the salient features of a variety of problems and puzzles in probability taken from a range of sources. Some problems are well known, some are new twists on old problems, and some have been a source of public debate. The selection is such that the working through each is important to develop a sound understanding of a different aspect of probability. Without necessary resort to symbolic mathematical expression or even calculation, eikosograms visually express the probabilistic structure of these problems and of the probabilistic relations and comparisons required for their solution. Eikosograms make probability, and probabilistic argument, much more natural.

A brief summary of some simple generic eikosograms introduces the display; it will become clearer and more familiar when used in the context of the problems. The full context of each problem is presented allowing the eikosograms embedded in situ in the descriptions and solutions.

The first problem shows the effect wording has on a person's ability to compare probabilistic outcomes. The data are based on a study by McNeil et al (1982) where people were given a life threatening medical scenario in which they had to choose a treatment for themselves from two available. Each treatment had different probabilities associated with categories of survival. Slightly different wording of the identical information produced very different choices. The eikosograms show how the same information would have been presented visually to facilitate more consistent comparison.

The next problem exercises the meaning of conditional probability by comparing strategies for choosing the cheapest gas station of three along a limited access highway. This is a novel context for a simple version of the so-called 'secretary problem' found in the statistical literature. The eikosogram belonging to the optimal strategy visually stands out as superior to the others considered.

The next turns again to medicine, now in the context of understanding the quality of medical tests – in particular the quality of a clinical assessment for osteoporosis recently proposed by Cadarette et al (2000). Eikosograms are used to introduce the concepts of the sensitivity and specificity of these tests. These eikosograms are put together via the prevalence of the condition being tested and show how joint probability distributions are constructed as the mixture of conditional distributions. The visual combination suggests using opacity = 1 - specificity as an equally natural alternative to specificity, one which more easily combines with sensitivity via prevalence. The variables are interchanged on the combined eikosogram and used to produce the dual eikosogram for the same information, now highlighting the true/false positive and the true/false negative rates of the test rather than its sensitivity and opacity. Besides underscoring important medical concepts and their inter-relationships, the transition from one eikosogram to the other is an important illustration of the Bayes rule connecting conditional probabilities.

A probabilistic understanding of uncertain tests is important in a large variety of situations unrelated to medicine (e.g. security, lie-detectors, drug testing, etc.). The next problem explores ‘tests’ in the context of jurisprudence, in particular the reliability of an eye witness to a hit and run accident involving a taxi cab. The problem draws together the ideas from the two previous medical examples – people’s understanding of probability, especially when expressed only verbally, and the Bayes relation. I review some classic studies which have shown that people making probability assessments of guilt or innocence not only do not follow Bayes rule but sometimes follow no consistent rule at all (e.g. Kahneman and Tversky, 1972). An important value of the eikosogram is that the visual display of the probabilities is simple, both to understand and to reason with, rendering it much more natural to follow the rules of probability.

Next considered is a problem which casts doubt on any claim that personal preferences are ordered. The context is an artificial one involving a person’s preference for pie and shows that it is perfectly rational for a person to prefer A to C, and B to C, yet also to prefer C to both A and B. It is also shown that a second person with the same information could equally rationally prefer A to C, and B to C, as did the first, yet when choosing between the three, prefer A. This demonstrates the heterogeneity of rational preferences across different people faced with identical information. Although the problem has been in use for years by different colleagues at Waterloo, and possibly by others elsewhere, the detailed context and the numerical values given here are original; so too is the extension of the problem by introduction of the second person.

In probability terms, the problem demonstrates the graphical determination of joint probabilities as well as the important visual structure of independent variables. Eikosograms are used to display and to visually compare various joint probabilities. Numerical values, as desired or required, are easily calculated from the eikosogram.

The previous problem’s solution is straightforward via eikosograms but is at the same time paradoxical in being at first counter-intuitive. Like all so-called paradoxes, it is much less puzzling in hindsight. The remaining problems have all at one time or another been called a paradox and, as such, have sometimes garnered considerable and public debate.

The most important of these paradoxes is Simpson’s paradox, an amalgamation paradox whereby the relation between two variables reverses when conditioned on a third. It is now standard treatment in introductory treatments of probability. Here an artificial, but plausible, context is presented involving four binary variables. The probabilities (original here) are plausible and constructed so that the Simpson’s paradox reverses twice. Eikosograms are shown to visually expose just how easy such reversals can be produced. This visual logic by eikosogram carries over to variables having any number of categories.

The next problem considered is usually called the ‘two envelope paradox’ and has been the subject of some debate, particularly in the philosophy literature but also by statisticians. Outcome trees and eikosograms show the paradox to be simply resolved. Perhaps, the most puzzling aspect of this problem is the amount of learned discussion it has generated.

The final problem discussed has a few variations. Two are classic and described here – one known as ‘the Monty Hall problem’ and the other (an earlier version) as ‘the prisoner’s dilemma’ or the ‘three prisoner problem’. The former erupted into debate in the popular press in the 1990s. The latter, according to Maynard Smith (1968, p. 70), “. . . should be called the Serbelloni problem since it nearly wrecked a conference on theoretical biology at the villa Serbelloni in the summer of 1966”. Indeed, M.S. Bartlett reviewed Maynard Smith (1968) for *Nature* where he discussed the problem and so sparked an exchange of letters to *Nature* with D.V. Lindley (Bartlett, 1970).

A difficulty with both problems, which contributes some to their continued discussion, is that they are often incompletely specified and different authors layer, sometimes only implicitly, further assumptions so that they might be resolved. Using outcome trees and eikosograms, their common structure and differences are revealed, as is the incompleteness of the information provided. Even with incomplete information, the eikosograms show that a reasonable strategy is available to the game show contestant in the Monty Hall problem and that a reasonable inference is available to the prisoner in the prisoner's dilemma.

The difference in probabilistic structure between the two problems is imposed by their different contexts. For each of these classic problems a new (and hopefully entertaining) variation is introduced which exploits this difference and presents new 'paradoxes'. Again, the solutions of these new, and more complicated, paradoxes is easily seen visually via eikosograms. Because of the increased complexity, the general treatment is supplemented by some simple mathematics derived entirely from the visual characteristics of the corresponding eikosgrams.

This illustrates how, even in relatively complex situations, eikosograms can be used to derive the mathematical expressions that help solve the problem. Symbolic and diagrammatic reasoning operate in parallel, strongly coupled via the eikosogram.

1.1 Section order

Section 2 is a brief synopsis of eikosograms, Section 3 the choice between medical treatments, Section 4 the problem of buying gasoline, Section 5 medical testing for osteoporosis, Section 6 the taxi cab problem, Section 7 the cherry pie problem, Section 8 Simpson's paradox, Section 9 the two envelope paradox, and Section 10 the Monty Hall problem, the prisoner's dilemma and a variation on each. Some brief concluding remarks are given in the final section.

2 Eikosograms

Figure 1 shows the eikosogram for a random variable, Z, which has exactly three different possible

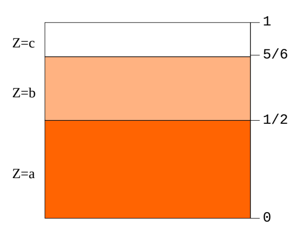
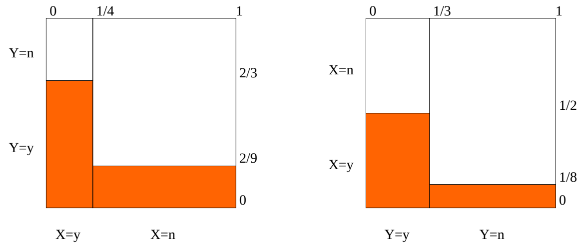


Figure 1: Eikosogram for the random variable Z taking values a, b, and c with respective probabilities $1/2$, $1/3$, and $1/6$.

values a, b, and c. Each value has an associated rectangle whose area is the probability that Z takes on that value. Because these rectangles have width one, the areas are determined entirely by the heights.

For two or more variables, one variable is associated with the vertical axis and the others with the horizontal axis. Figure 2 shows the joint distribution of two binary variables Y and X (each taking values y and n, for 'yes' and 'no').



(a) $Y|X$ format (b) $X|Y$ format

Figure 2: Joint distribution of two binary random variables Y and X.

In Figure 2(a), the width of the vertical strips is determined by the marginal probabilities, $1/4$ and $3/4$, that X takes values y and n respectively. The height of the shaded rectangle within each vertical strip is the conditional probability, $2/3$ and $2/9$, that $Y = \mathsf{y}$ given $X = \mathsf{y}$ and given $X = \mathsf{n}$ respectively; the height of unshaded rectangles correspond to the conditional probabilities, $1/3$ and $7/9$, that $Y = \mathsf{n}$ given each value of X. The areas of each rectangle is the joint probability for the corresponding values of Y and X – e.g. $Pr(Y = \mathsf{y}, X = \mathsf{y}) = 1/6$.

Because there are two variables, the distribution could alternatively have been expressed with X on the vertical axis and Y as the conditioning variable on the horizontal axis as in Figure 2(b). Matching conditions and hence areas is a visual statement of Bayes relation between conditional probabilities – e.g. area of the lower left rectangle in both eikosograms is the same, namely $Pr(Y = \mathsf{y}, X = \mathsf{y})$ which, calculating $height \times width$, is $Pr(Y = \mathsf{y}|X = \mathsf{y}) \times Pr(X = \mathsf{y})$ in Figure 2(a) and $Pr(X = \mathsf{y}|Y = \mathsf{y}) \times Pr(Y = \mathsf{y})$ in Figure 2(b).

See Cherry and Oldford (2003) for further detail on how axioms for probability are motivated by and coincident with these pictures – so too are the concepts of probabilistic independence and conditional independence. The entire independence structure of three variables (or three groups of variables) is exhaustively explored in Oldford (2003), including summary by log-linear and graphical models and the development of graph-based theorems derived from features of eikosograms.

3 Choosing between medical treatments

A patient has been diagnosed to have a disease which, unless treated immediately, is certain to take her life. There are two possible treatments which have been used with varying success in the past. To arrive at an informed decision, the patient and physician discuss the various outcomes and the known chances with which they might occur. Ultimately, the patient chooses one or the other treatment. Both patient and physician want this decision to be based on the information available and not on the manner of its presentation.

McNeil et al (1982) investigated a hypothetical situation where the disease was lung-cancer and the patient had to choose between either surgery or radiation treatment. They imagined the physician presenting the possible outcomes and their associated chance of occurring as follows (reproduced here as in Tversky and Kahneman, 1988):

Surgery: Of 100 people having surgery 90 live through the post-operative period, 68 are alive

> at the end of the first year, and 34 are alive at the end of five years.
>
> *Radiation therapy*: Of 100 people having radiation therapy, all live through the treatment, 77 are alive at the end of one year, and 22 are alive at the end of five years.

When 247 undergraduate students were asked to put themselves in the position of the patient, McNeil et al (1982) found that 18% said they would choose radiation treatment.

However, the same statistical information could just have easily been presented as follows:

> *Surgery*: Of 100 people having surgery 10 die during surgery or the post-operative period, 32 die by the end of the first year, and 66 die by the end of five years.
>
> *Radiation therapy*: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year, and 78 die by the end of five years.

Presented in this way, to a similar group of 336 undergraduates, the proportion who chose radiation thereapy more than doubled to 44%!

In the first case, the outcomes of the options were framed entirely in terms of survival and surgery was much more appealing than radiation. When framed in terms of mortality, as in the second case, the possibility of death during surgery seems to have been heightened and consequently reduced surgery's appeal for more people. The statistical information is identical in both cases, but its perception, and hence the decision, was affected by the language in which it was framed.

In a repeat study (percentages found this time were 18% and 47%), McNeil et al (1988) introduced a third presentation of the statistical information framed by mortality and survival together as follows:

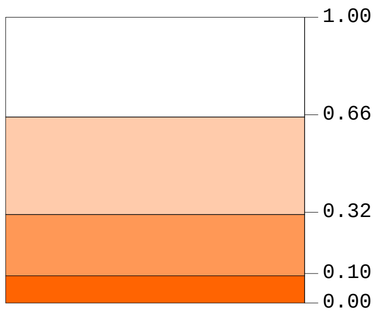
> *Surgery*: Of 100 people having surgery 10 will die during treatment and 90 will live through the treatment. A total of 32 people will have died by the end of the first year and 68 people will be alive at the end of the first year. A total of 66 people will have died by the end of 5 years and 34 people will be alive at the end of 5 years.
>
> *Radiation therapy*: Of 100 people having radiation therapy, none will die during treatment (i.e. all will live through the treatment). A total of 23 people will have died by the end of the first year and 77 people will be alive at the end of the first year. A total of 78 people will have died by the end of 5 years and 22 people will be alive at the end of 5 years.

They found that, with this framing, 40% of the participants chose radiation treatment over surgery, closer to the results found for the mortality frame.[1]

A difficulty common to all three frameworks is that the statistical information is extracted with difficulty, being embedded in text (separate for each treatment) and presented only in terms of cumulative probabilities (e.g. 68 out of 100 live at least one year after surgical treatment). Granted this is better than would be a presentation received only aurally (as is often the case in reality), but a visual arrangement of the information would permit further separation of the statistical and emotive content.

Figure 3 shows the statistical information displayed as eikosograms. The information in the

[1] Note that in each case the mortality figure preceded the survival figure; the reverse was not investigated.



(a) Surgery (b) Radiation

Figure 3: 200 patients diagnosed with lung-cancer – 100 receive (a) surgery, 100 (b) radiation treatment. From bottom to top the categories are y_1= "Does not survive treatment", y_2 = "Survives treatment, but only to one year, y_3= "Survives more than one but fewer than five years" and y_4 = "Survives at least 5 years". The area (or equivalently the height) of each shaded rectangle matches the proportion of the 100 which are in that category. The shading matches the category across the two figures and for radiation the bottom most category, y_1, is absent because all survive radiation treatment.

caption accompanying the figures isolates the emotive context. These eikosograms facilitate and encourage visual comparison between the two treatments without restricting focus to any particular feature. Treatments are compared by category by comparing areas having the same shade in the two eikosograms. Cumulative probabilities are determined and compared either by accumulating area top down (survival probabilities) or bottom up (mortality probabilities).

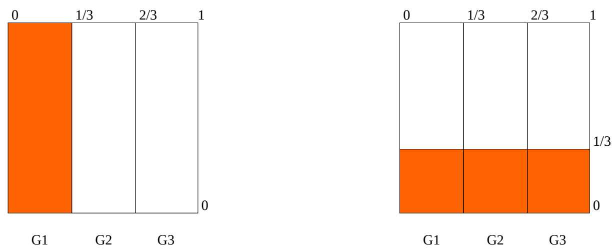
While decision on treatment could still be influenced by framing the eikosogram either entirely in terms of mortality, or entirely in terms of survival, presenting the information more even handedly should be facilitated by this visual presentation. The even handedness would be further enhanced by a second vertical axis on the left of each eikosogram to mark cumulative survival probabilities top down from zero to one. It would be interesting to repeat the McNeil et al (1982, 1988) framing experiments with eikosograms and accompanying text as in Figure 3.

4 Buying gasoline

Suppose that you regularly commute along a limited access divided highway and that three highway gas stations are located along the route. Each gas station advertises the price of its gasoline on a sign just at the beginning of the off ramp leading to the station. Drivers must decide at that time whether to exit the highway to purchase gasoline or to continue on to the next station; once past a gas station there is no turning around and going back.

Suppose further that the price of gasoline is different at all three stations and that the station with the lowest price could be any one of the three. Your experience shows that no one of the three gas stations has the lowest price any more often than the other two. Moreover, gas prices fluctuate enough that it is difficult to tell whether on any given day a particular price is low or high without actually comparing it to the others. All we know then is that, for any given trip, the probability that any one of the three gas stations – G1, G2, or G3 – is cheapest of the three is the same, namely 1/3. Being a regular commuter, your objective is to pay the least amount for gas that you can.

Two different strategies immediately suggest themselves. First, always choose the same gas station, e.g. G1. At least then you will have the cheapest gas one third of the time. Of course, you also have the most expensive one third of the time. The effect of this strategy is illustrated in Figure 4(a). There, the horizontal condition G1 means that G1 has the cheapest gas, G2 that G2



(a) Always buy at G1 (b) Choose gas stations equiprobably

Figure 4: Horizontal axis identifies that gas station as cheapest; vertical axis the probability of buying gas there given it is cheapest. Total shaded area is the probability of having paid the least with that strategy.

has the cheapest, and G3 that G3 does. The vertical probability is the probability that, by following this strategy, you have purchased the cheapest given the condition – when G1 is the cheapest, you will buy gas there with probability 1; you never buy at G2 or G3, so whichever is cheapest the probability of purchase there is zero. The probability of getting the cheapest gas is determined from the shaded region which, as expected, has area 1/3. Changing the role of the horizontal axis to be that of the most expensive gas station and the same eikosogram will result, showing the probability of purchasing the most expensive gas to be 1/3 as well.

The second strategy is to choose a gas station at random. Figure 4 illustrates the result of this strategy when the gas stations are chosen with equal probability. As can be seen, this has not changed the probability of getting the cheapest gasoline – it remains one third. This probability would not change if the gas stations were selected with different probabilities – the heights of individuals bars would change but the total shaded area would still be 1/3.

With these two strategies, it is as if one can do no better than to move the total probability of 1/3 around the eikosogram. However, neither strategy has made use of the gas prices already seen. With the exception of the first station, whenever we approach a new station we know the price at that station and at all stations we have passed without purchasing.

This suggests a third strategy. Always pass G1 but observe the price of its gasoline. If G2 should have cheaper gas than G1, then purchase gas at G2; otherwise purchase gasoline at the third station, G3. The probabilities associated with this strategy are shown in Figure 5. There we see, even if G1 is the cheapest, its gasoline will never be purchased by us. If G2 is cheapest, then its price will be less than G1's and its gas will certainly be purchased. If G3 is cheapest, then G1's price will be greater than G2's half the time and less than G2's half of the time – when it is less we will buy from G3, when it is more, from G2. By following this strategy, the total probability that we buy the cheapest gas is the total area of the shaded regions, which is $1/3 + (1/3 \times 1/2) = 1/2$.

Comparison of the shaded areas in Figure 5 and Figures 4(a) or (b) show that this strategy dominates each of the other two. In spite of there being three gas stations, the strategy ensures that

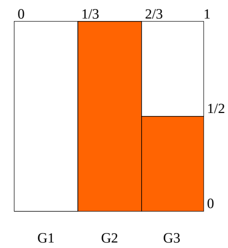


Figure 5: Making use of the known prices.

half the time we purchase from the cheapest. Moreover, this improvement has been at the expense of the most expensive gas station – a similar exercise shows that the most expensive gasoline will be purchased only 1/6 of the time by following this strategy.

This problem is a special case of a more general problem of optimal stopping, the 'secretary problem', traditionally cast in terms of hiring a secretary. Here candidates for the position are interviewed in sequence and hiring must occur immediately after a person's interview or not at all. As with the gas stations, once the decision to interview the next candidate is taken there is no going back to the previous candidates. The objective is to have a strategy which maximizes the probability of hiring the best person for the job.

5 Medical testing

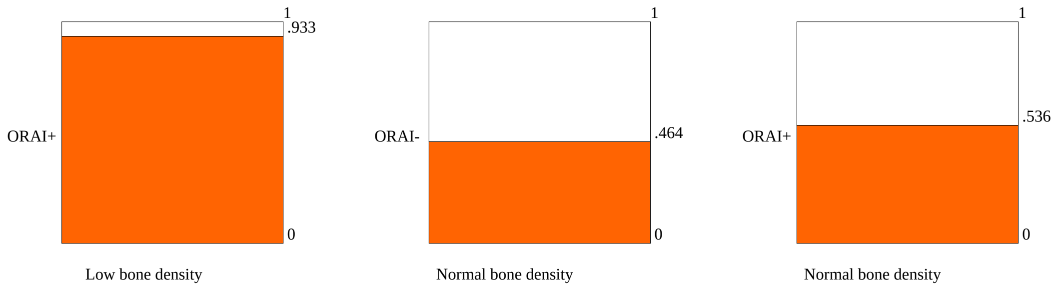
In medical diagnosis, tests are often administered to determine whether a patient does or does not have some specified condition (e.g. pregnancy, bacterial infection, presence of cancer cells, etc.). Tests tailored to the condition return a 'positive' result if the test determines presence of the condition, and a 'negative' result otherwise, indicating absence of the condition. Of course testing positive, or negative, does not necessarily imply that the condition is indeed present, or absent; it could be that the test is mistaken.

Ideally, a test has high probability of being 'positive' when the condition is present and a high probability of being 'negative' when the condition is absent. The first probability is called the *sensitivity* of the test, the second its *specificity*. A test which has both higher sensitivity and higher specificity than another is preferred, all other things being equal (e.g. side-effects, costs, etc.), over the other. However, sensitivity and specificity only compare tests; they are not enough to determine the probability that a patient who tests positive actually has the condition. For that, we also need to know the proportion of people who have the condition in the population, called the condition's *prevalence*.

Osteoporosis is a disease whereby one's bones become less dense over time, and hence more easily broken and compressed, and can eventually lead to skeletal failures. The disease is particularly prevalent in post-menopausal women, affecting about 1 in 4 Canadian women over 50 (Osteoporosis Society of Canada, 2003), but also affects men (about 1 in 8 Canadian men over 50).

A recent statistical study by Cadarette et al (2000) developed a new clinical assessment tool, which they called ORAI (for Osteoporosis Risk Assessment Instrument), to help physicians determine which women should undergo more extensive (and more expensive) testing of their bone mineral density (BMD) for osteoporosis. In their assessment of ORAI they estimated its specificity to be 93.3% and its specificity to be 46.4% for selecting women with low bone mineral density.

In Figure 6(a) the sensitivity is the height of the shaded rectangle as determined by the condi-



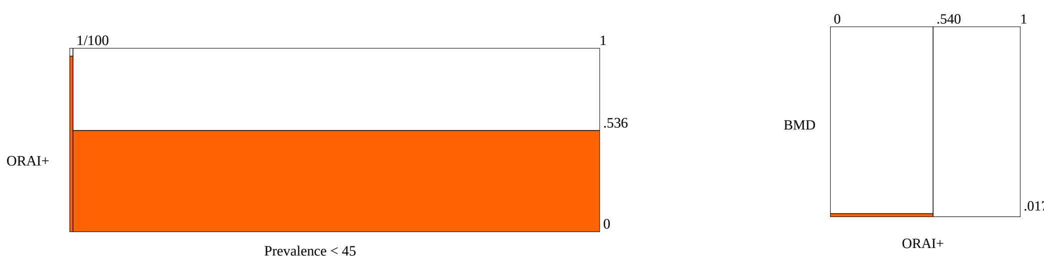
(a) Sensitivity (b) Specificity (c) Opacity = 1- Specificity

Figure 6: Assessing the value of the ORAI test for low bone mineral density (BMD).

tional probability of a *positive* ORAI test given that the patient has low BMD. Figure 6 (b) shows the specificity as the height of the shaded rectangle which here is the conditional probability of a *negative* test given the patient does *not* have low BMD. Figure 6(c) is the same as Figure 6(b) except that the height of the shaded region is now one minus that of Figure 6(b) or the conditional probability of a *positive* test given the patient does *not* have low BMD. This probability, 1-specificity, might reasonably be called the *opacity* of the test – low opacity being the same as high specificity.

The advantage to casting specificity in terms of opacity is that the shaded areas in Figures 6(a) and (c) each correspond to the conditional probability of testing positive, the first given low BMD and the second given no low BMD. This means that the two eikosograms can be merged into one when the relative prevalence of the two conditions is known.

For example, Caderette et al (2000) estimated that the prevalence of osteoporosis in Canadian women under 45 years of age is less than one per cent. In Figure 7(a) the sensitivity and opacity



(a) Prevalence for women under 45 (b) Conditional on test

Figure 7: Assessing the value of the ORAI test for low bone mineral density.

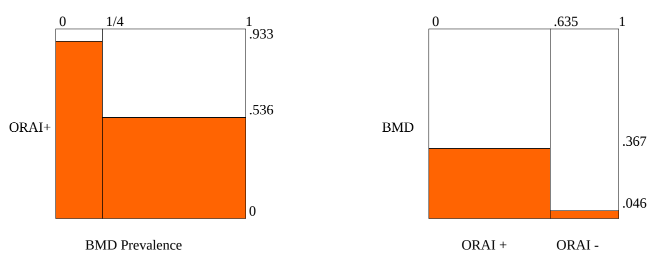
eikosograms of Figure 6 (a) and (c) are combined as if the prevalence of low BMD in these women is exactly one per cent. The left most strip is the sensitivity eikosogram, now having width 1/100, and the right most strip of width 99/100 is the opacity strip. The horizontal axis has stretched so that the 1% strip can be seen; although no longer a square, all probability calculations are as before.

Incorporating the prevalence shows that the vast majority of the women in this population who would test positive (area of entire region shaded), would not in fact have low BMD (proportion of the entire shaded area belonging to the larger rectangle). Routinely applying ORAI to this test would therefore turn up a great many *false positives*; on the other hand, visual inspection also indicates that few *false negatives* would show up. This perspective on the information is more easily seen after interchanging the horizontal with the vertical variables and reconstructing the eikosogram. The result is shown as Figure 7(b).

In Figure 7(b), the total shaded region now marks the proportion of women who have low bone mineral density and, as before, is 1%. Most of this 1% is where it should be, in the left most strip where the women have tested positive. Unfortunately, the strip is very wide and so only about 1.7% of these women will actually have low BMD; the rest, 98.3%, will be false positives. The corresponding probability of having low BMD given a negative test is much smaller still and does not even show up in the display. This is good news, for the height of the right shaded bar is the false negative rate for ORAI and is seen to be exceedingly low. This extremely low rate means that routine application of ORAI to the under 45 population of Canadian women would rule out 46% ($= 1 - 0.540$) of them as having low BMD.

As a matter of public policy, the low probability of a *true positive* and the age progressive nature of osteoporosis together suggest that routine testing of women under 45 may be a poor use of public health resources. Caderette et al (2000) recommend against routinely testing women under 45.

More reasonable would be to consider Canadian women over 50 where the prevalence of both osteoporosis and low BMD is much higher. Taking these prevalences (i.e. low BMD and osteoporosis) to be the same in this population (1 in 4), the corresponding eikosograms are given in Figure 8. As can be seen in Figure 8(a), the sensitivity and opacity are as before but the greater



(a) Prevalence for women over 50 (b) Conditional on test

Figure 8: Prevalence in Canadian women over 50.

prevalence of low BMD in this population (1/4 rather than 1/100) has produced a substantial effect on the probabilities in Figure 8(b).

Routinely testing this population by ORAI is more easily justified. Fully 36.5% will test negative and of these only 4.5% would be false negatives. Those who do test ORAI positive (63.5%) still have a relatively high false positive rate of 63.3% ($= 1 - 0.367$) and should be tested further by measurement of their bone mineral density, a more costly procedure called bone mineral densitometry. In the light of rising health care costs, Caderette et al (2000) recommend this two step procedure of first screening all women over 50 with the simple ORAI test and then measuring the BMD only for those who tested positive.

Similar testing situations arise outside of medical practice. For example, in recent years some employers have decided to routinely test employees and/or applicants for illicit drug use. Convinced by assessment of the test's sensitivity and specificity, employment decisions have been made based entirely on whether the person tested positive or negative. Yet the only probability on which such decisions could reasonably be based is the conditional probability that the person is a drug user given they test positive, and this cannot be determined without information on the prevalence of drug use in the group being tested.

Whatever the testing situation, the same reasoning, via eikosograms like those of Figure 8, applies.

6 Green cabs, Blue cabs

"A cab was involved in a hit and run accident at night. Two cab companies, the Green and Blue, operate in the city. You are given the following data:

(a) 85% of the cabs in the city are Green and 15% are Blue.

(b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green?"

This problem was first introduced by Kahneman and Tversky (1972) and has been presented to many subjects in slight variations by different investigators. The above version is that given in Tversky and Kahneman (1982, pp. 156-157). There, they report that the median and modal answers given by these subjects is typically 0.80, exactly the probability that the witness correctly identifies the cab colour in the test situations.

Knowing that the witness has been demonstrated to be right 80% of the time, both when the cab is Green and when the cab is Blue, it would seem to many of these potential jurors that the cab has probability of 0.80 of being whatever colour the witness says it was. The distribution of the cabs in the city seems to have been deemed irrelevant.

That this is not the case might have been more easily seen had the information been presented visually. The eikosogram which records the information as given in the problem is shown in Figure 9. The Green cab, Blue cab distribution determines the vertical widths, the witness's reliability the height of the shaded bars within each strip. The height of each vertical bar is the probability that the witness identifies the cab as Blue either given it is Green (left vertical strip) or given it is Blue (right vertical strip).

The probability of evidentiary interest is

$$Pr(\text{Cab is Blue} \mid \text{Witness says it is Blue})$$

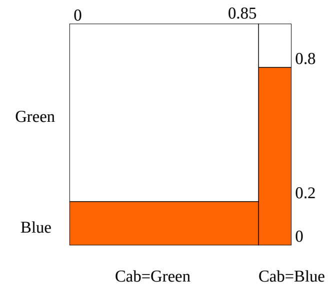


Figure 9: The Green cab, Blue cab problem as given.

and not

$$Pr(\text{Witness says it is Blue} \mid \text{Cab is Blue}).$$

The latter, like sensitivity in the medical testing example, only says something about the witness's reliability in identifying Blue cabs. It is as if the subjects confused these two probabilities.

From the eikosgram of Figure 9, the first probability is seen to be the ratio of the area of the right most shaded region to the entire shaded region. Immediate visually, this ratio is much less than 0.80, possibly less than 1/2; calculation of the ratio of areas reveals it to be 0.414. Thus the probability that the cab is Blue, given the witness says it is, is much less than the reliability, 0.80, the witness.

The full evidentiary story is made more immediate, as with the medical testing example, by interchanging the variables on the eikosogram, all the while preserving areas and hence probabilities. The eikosogram from this perspective is shown in Figure 10. Here the height of the shaded

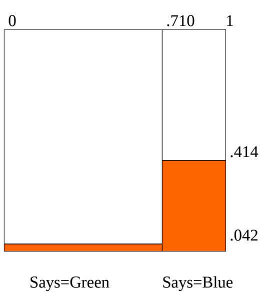


Figure 10: Cab colour conditional on witness takes account of both witness reliability and base rates.

bars is the probability that the cab is Blue given the witness either says it is Green (at left) or says it is Blue (at right). The total of the shaded area is the probability that the cab is Blue unconditionally and must equal the area of the same in the previous eikosogram of Figure 9, namely 0.15.

As can be seen from either eikosogram, the witness saying the cab is Blue provides little evidence one way or the other. However, were the witness to have said the cab was Green, the evidence that the cab was indeed Green would have been quite high – a probability of 0.958 ($= 1 - 0.042$).

6.1 Why was the wrong probability used?

In other studies, the subjects were not presented the witness information in (b) but only the distribution of Green and Blue cabs as in (a). In this case, almost all subjects gave the base rate, 15%, as the probability that the cab was Blue. Being the marginal probability (i.e. no witness), this answer is correct. With the witness present, the subjects simply ignored the base rates of Green and Blue cabs.

Bar-Hillel (1980) replaced the witness identification by a report that 80% of Blue cabs have an intercom installed, as do 20% of Green cabs; a witness, who did not see the cab, heard its intercom as it passed in the night. Faced with this information, the median answer produced was 0.48. Although the median was much closer to the correct answer of 0.414, the mode was 0.30 and the answers were more variable than in other cases. The variability suggests that no uniform strategy was employed to combine the probabilities from the two sources.

Even so, unlike the original problem, the probabilities were somehow combined. Bar-Hillel (1980) suggests that the reason for this is that the intercom information appears to be no more specific than the incidental base rate of the two cab companies, and so the two sources are perceived to have roughly equal merit. In the original problem, the witness specifically identified the cab company and the perception was that this rendered he base rates irrelevant.

Alternatively, it might be argued that the subjects in the original studies did not confuse the two probabilities, but rather thought that they had no reason to suppose that one cab company was just as likely to be involved in an accident as the other. Although the Green cabs are more plentiful than the Blue cabs, perhaps the Blue cabs are involved in more accidents than the Green cabs, or vice versa. Not knowing which of these was the case, the subjects clung to the one probability they had, the reliability of the witness.

To assess this, in some studies the information in (a) was replaced by

(a') Although the two companies are roughly equal in size, 85% of cab accidents in the city involve Green cabs and 15% involve Blue cabs.

Subjects adjusted their assessment of the probability down in light of this causal information; such studies produced median answers of about 0.60 compared to the correct answer of 0.414. Tversky and Kahneman (1982) explain that, when the base rate is of a causal rather than incidental nature, then it is more likely to be perceived to have value.

Two problems have emerged. First, it would seem that people choose *when* to combine base rate information with other evidence based on the perceived relative relevance of the information and not on its probability structure. The above studies have shown that base rate information is used when it is all that is available, or in conjunction with other information either when the base rate has causal content or when it is no less specific than the second source. Otherwise, the base rate information may not be used at all. Second, when the two sources are both deemed relevant, *how* the combination is done does not follow the rules of probability and is neither uniform across subjects nor across presentations of the information.

Since such judgements are routinely required in courts of law and elsewhere, it is important that the probability structure be understood and the rules of probability followed. As was the case with choosing between medical treatments, it may be that the problem lies to some extent with the textual presentation of the information and that a visual presentation via eikosograms would more often have people produce the correct answer.

7 Cherry pie

A local cafe serves at most three kinds of pie – apple, blueberry and cherry. A restaurant critic once visited the cafe and wrote a rave review which described the cherry pie he had ordered as superb. Since then, the cafe has made the cherry pie a permanent part of its menu. One or the other of apple and blueberry are also offered daily but only occasionally are all three offered. Dan, a regular customer at the cafe, always orders a coffee and a piece of pie. Recently, his choice of pie surprised the servers at the cafe.

Every time Dan has had to choose between apple and cherry, he has invariably chosen apple. And every time the choice is between blueberry and cherry, Dan always chooses blueberry. Observing this pattern, the servers would know to bring Dan whichever other pie was being offered that day. The clear conclusion was that Dan did not care for cherry pie, however good it might be.

And so, on the first occasion that all three varieties were available, the server did not know Dan's preference and had to ask which kind of pie he preferred: apple or blueberry. The server was understandably taken aback when Dan responded "Neither, I prefer cherry please."

The question is whether Dan is behaving rationally. When presented with apple and cherry, he prefers apple; when presented with blueberry and cherry, he prefers blueberry. But when all three are available, his preference for cherry seems inconsistent.

In fact, Dan's behaviour turns out not only to be consistent and rational but, in a sense, optimal.

7.1 Pies of differing quality

Sad to say, but the quality of the pies served at the cafe, as Dan knows only too well, can vary from day to day. Measuring the quality of pies on a 6 point scale, where 1 is the highest quality and 6 the lowest, the apple pie is invariably a good solid quality 3 pie. The blueberry pie is quite variable in quality: half the time it is excellent at 2, 45% of the time it is a 4 and 1 in 20 times it is awful having the lowest possible quality, 6. The cherry pie can be extraordinary achieving the highest quality of 1 fully 40 % of the time but the rest of the time it is a poor quality 5 pie – the restaurant critic clearly came on a good day for cherry pie.

Table 1 summarizes the proportion of pies at each quality for the three types of pie. On any given day that blueberry is served, it will be a quality 2 pie with probability 0.50, quality 4 with probability 0.45, and quality 6 with probability 0.05. Similarly, the probabilty that the cherry pie served will be a quality 1 pie is 0.40 and that it will be a quality 5 pie is 0.60 . Moreover, the quality of blueberry pie is independent of the quality of cherry pie – e.g. the probability that a blueberry pie is of quality 2 remains 0.50, whatever the quality of the cherry pie served.

Apple		Blueberry		Cherry	
Quality	Proportion	Quality	Proportion	Quality	Proportion
3	1.00	2	0.50	1	0.40
		4	0.45	5	0.60
		6	0.05		

Table 1: Pies of varying quality served at the cafe. Highest quality is 1; the lowest is 6.

The joint probability of the qualities of blueberry and cherry pies, can now be displayed in an eikosogram as in Figure 11. Here blueberry's quality, B, appears on the vertical axis and cherry's,

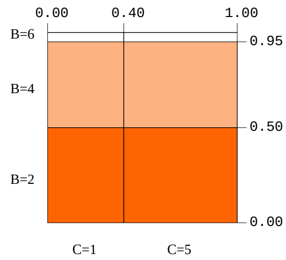
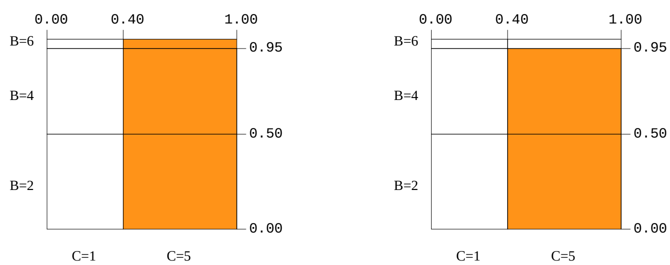


Figure 11: Cherry Pie: Quality high to low is 1 to 6; B is blueberry, C is cherry, and apple is always of quality 3. Shading here reflects the different quality values for blueberry pie. The flat structure follows from the independence of the quality of blueberry and cherry pies.

C on the horizontal; shading reflects the values of B. In fact, because apple always has quality $A = 3$, this eikosogram is also the joint distribution of the qualities of all three pies. The area of the lower left rectangle, for example, is the probability that all three pies are at their best – apple pie is of quality 3, blueberry of quality 2, and cherry of quality 1 – or $Pr(A = 3, B = 2, C = 1) = 0.50 \times 0.40 = 0.20$.

This eikosogram can be used to reveal Dan's reasoning. In Figure 12(a), the shaded region

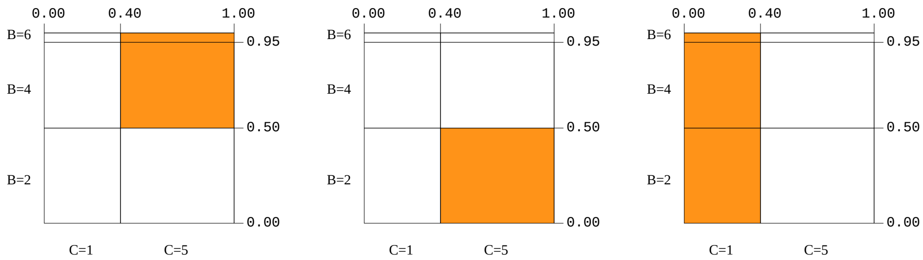


(a) Apple better than Cherry (b) Blueberry better than Cherry

Figure 12: Shaded areas are the probabilities that each pie is of better quality than cherry.

identifies where apple is better than cherry (i.e. $A < C$) showing it to have area, and hence probability, 0.60 – larger than the probability that cherry is better. When the choice is between these two pies, Dan has higher probability of getting the better pie by choosing apple. Similarly, when the choice is between blueberry and cherry, the probability that blueberry is better than cherry (i.e. $B < C$) is the area of the shaded region of Figure 12 and, at 0.57, can be seen to be larger than the unshaded area. Again, Dan has higher probability of getting the better pie by not choosing cherry.

When all three pies are available, Dan's strategy is the same – to have the best pie with the highest probability. Figures 13 (a), (b), and (c) show, respectively, the regions where apple, blue-



(a) Apple is best (b) Blueberry is best (c) Cherry is best

Figure 13: Shaded areas are the probabilities that each pie is better than both of the others.

berry, and cherry are each the best of the three. By inspection, the shaded areas of Figures 13(a) and (b) are identical (both equal 0.30) and less than the shaded area (0.40) of Figure 13(c). Cherry, now having the highest probability of being the best pie of the three, will be Dan's choice when the all three are available.

Dan's reasoning is consistent and optimizes the probability that he will get the best pie available on each occasion.

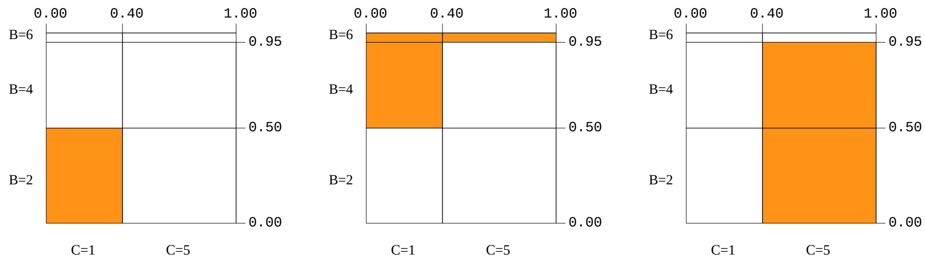
7.2 Enter Emma

Emma, a friend of Dan, also goes regularly to the cafe for coffee and pie and, like Dan, she never chooses cherry when only two different types of pie are available. She too is a careful reasoner and familiar with probability. Emma and Dan were sharing a table on the occasion that all three types of pie were available. Having puzzled the server, Dan carefully explained his reasoning for now choosing cherry pie. The server seemed satisfied and, turning to Emma, asked whether she too would now have a piece of cherry pie. Emma stunned everyone by saying that, although she loves cherries, it would be best for her to have the apple pie.

Emma's reasoning is consistent and identical to Dan's with one important difference. Rather than maximizing the probability that she gets the best pie available, Emma prefers to minimize the probability that she gets the worst. In the eikosograms of Figure 12, the shaded areas show Emma the probability that cherry is the worse of the two pies available and clearly indicate cherry is to be avoided. When all three pies are available, the probability that each is the worst of the three is given by the shaded areas of Figures 14. (a), (b) and (c). Clearly apple has the lowest probability of being the worst selection and cherry the highest. Emma is justified in choosing apple over the other two.

8 Good for each, bad for all

Suppose two medical treatments, A and B, have been developed to treat a fatal disease. A study is done to see which is the more effective: of 320 people who have contracted the disease, half are given treatment A, half treatment B. Of those who receive treatment A, 60 % ultimately survive



(a) Apple is worst (b) Blueberry is worst (c) Cherry is worst

Figure 14: Shaded areas are the probabilities that each pie is worse than both of the others.

the illness and 40 % do not; of those given treatment B, half survive and half die. Clearly A is the more effective of the two treatments.

During the study it was noticed that women more often survived than did men. This perception was later checked against the data and it was found that, whichever treatment they received, the percentage of women who survived the disease was indeed higher than the percentage of men who received the treatment.

Such gender-based differences are of considerable interest and so the data for women was separated and examined; half of the patients were women and their results appear in Table 2. As

	Treatment A	Treatment B
Die	39	11
Live	81	29
% Live	67.5	72.5

Table 2: Survival data for the 160 women of the hypothetical study.

can be seen, not only is the women's survival rate higher than the overall survival rate for each treatment, but the preferred treatment for women has changed! For women, the recommendation must be that B is the more effective treatment, not A.

Pleased to have discovered this important gender difference, the investigators then turned their attention to the results for the men. These data are shown in Table 3. As expected, for both

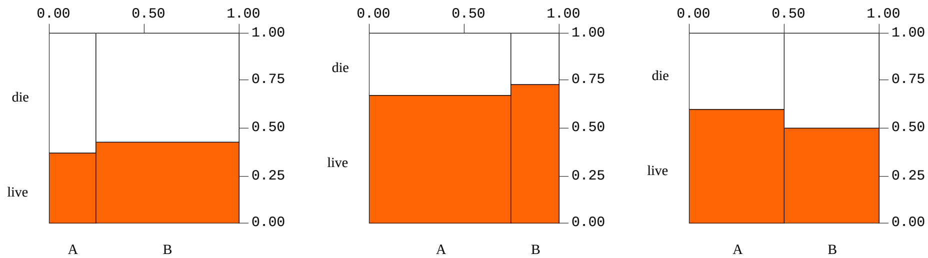
	Treatment A	Treatment B
Die	25	69
Live	15	51
% Live	37.5	42.5

Table 3: Survival data for the 160 men of the hypothetical study.

treatments the survival rates for men were lower than the overall rates. What was unexpected was that here too the more effective treatment is treatment B, not treatment A!

The investigators seem to have results which show treatment B to be better than A for women, and for men, but not for people.[2] This puzzling reversal in the relationship between treatment and outcome is called “Simpson’s paradox”.

The paradox and the conditions which create it are easily seen via the eikosograms for each sex separately and for both combined as shown in Figures 15(a)-(c). The puzzle is that the relative



(a) Good for men (b) Good for women (c) Bad for people

Figure 15: Simpson’s paradox: Hypothetical data on survival under two different treatments, A and B.

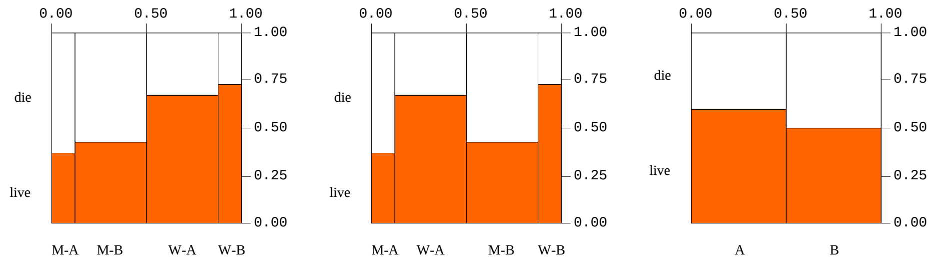
heights of the shaded bars in the two separate eikosograms switch when the data are amalgamated; the explanation is that the relative widths of the same bars are quite different in the two separate eikosograms – with probability, it is area (i.e. $width \times height$) not height alone, which is combined by amalgamation.

Figure 15 (a) clearly shows that most of the men were treated with B; moreover, most men did not fare well whichever treatment they received. Similarly, Figure 15 (b) shows that most of the women received A and that women had a relatively high chance of survival with either treatment. When the data are pooled by ignoring sex, the higher survival rate of women pulls up the performance of A because so many women received A and, simultaneously, the lower survival rate of men pulls down the performance of B. The result is as shown in Figure 15 (c).

More precisely, we begin with the joint distribution of three variables – survival, treatment and sex – and then stop distinguishing between the sexes to produce the joint distribution of survival and treatment. The first distribution is easily constructed from Figures 15 (a) and (b) by adjusting the width of each eikosogram to be the proportion of patients in that group (here this is 1/2 since there are equal numbers of men and women in the study) and then putting them side by side as shown in Figure 16(a). The vertical bars have been labelled with M or W, to identify sex, and A or B, to identify treatement. Figure 16(b) has the vertical bars re-arranged so that the treatment A is entirely on the left and treatment B on the right. Such re-arrangement cannot affect the joint distribution (all matching regions have identical area in either arrangement) nor the paradox (e.g. comparing the height of the shaded region of M-A to that of M-B still shows treatment B to have higher survival than treatment A for men and comparing W-A to W-B shows the same for women).

The advantage of the re-arrangement is twofold. First, it facilitates comparison of area between the eikosograms of Figures 16(b) and (c) – e.g. the two shaded rectangles at the left of Figure 16(b)

[2] This turn of phrase and the above context, though not the numbers, come from Baker and Kramer (2001).



(a) Men then women (b) A then B (c) No sex distinction

Figure 16: Water container metaphor: (a) Each vertical strip is a container of water. To combine the probabilities over the sex variable, first arrange that the containers having the same treatment are side by side (b), then remove the barrier separating the sexes within each treatment and allow the water levels to settle as in (c). The result is the joint distribution of survival and treatment (c).

must have total area equal to the area of the single shaded rectangle at the left of Figure 16(c). Second, the re-arrangement permits a simple metaphor for pooling the information across sex.

Imagine the vertical strips of Figure 16(b) to be containers of water (the shaded areas). The container W-A has a lower level of water than does W-B; similarly, M-A has a lower level than does M-B. To no longer distinguish between sexes, both the barrier separating containers M-A and W-A and the barrier separating M-B and W-B are simply removed and the water allowed to settle to new levels in the combined containers. The results are the water levels seen in the containers A and B of Figure 16(c) (total probability, like the volume of water, is preserved by this marginalization). Because the volume of the water in the high-level W-A is relatively large while that in the high-level W-B is relatively small, when put together the level of water in container A is higher than that in B. That this can happen should be no surprise.

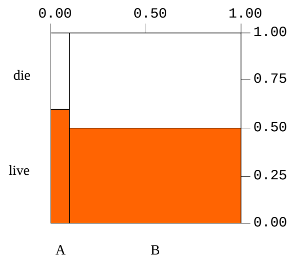
The reversal in the relative performance of the two treatments is simply a consequence of the joint probability of all three random variables with its imbalance between the proportion of men receiving treatment A (or B) and the proportion of women receiving A (or B).

Of course the investigators need not stop at distinguishing the patients by gender. They might, for example, consider separating each group into two age categories – younger and older – with results as shown in Table 4. The corresponding eikosograms are shown in Figure 17.

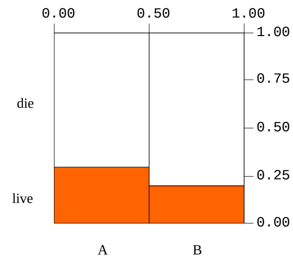
	Treatment A				Treatment B			
	Men		Women		Men		Women	
	Older	Younger	Older	Younger	Older	Younger	Older	Younger
Die	21	4	36	3	24	45	5	6
Live	9	6	54	27	6	45	5	24

Table 4: Hypothetical data for a twice reversing Simpson's paradox.

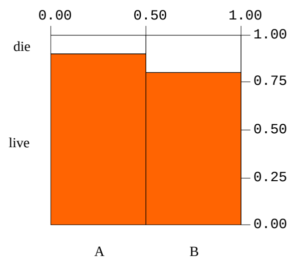
As might be expected, younger men and younger women both survive better than do their older



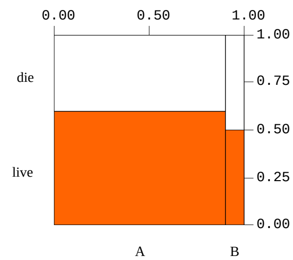
(a) Younger Men



(b) Older Men



(c) Younger women



(d) Older women

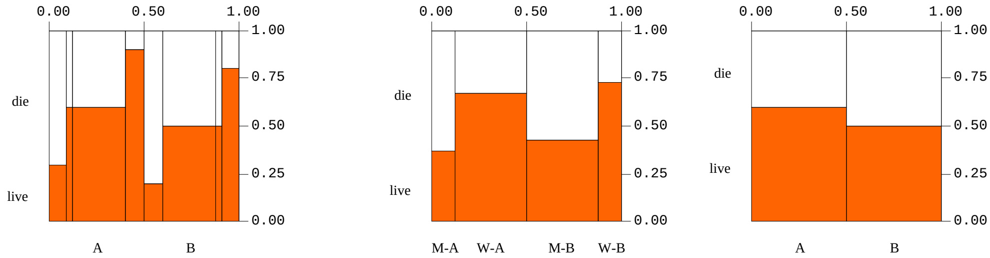
Figure 17: Eikosograms for all of the data disaggregated by both sex and age categories.

counterparts. What is unlikely to be expected is that treatment preference has again reversed! In every case – younger women, younger men, older women, older men – A is now preferred to B!

Again, as easily seen in the eikosograms of Figure 17, there is an imbalance in the portioning of the treatments to each younger and older group within sex. That the double reversal is due to this imbalance can be seen via the water container metaphor and the eikosograms of Figure 18.

Having gone from preferring A to B for all, to preferring B to A for men and for women separately, to again preferring A to B for younger and for older women and for younger and for older men, it would not be surprising if the investigators were at a loss as to what their conclusion should be! A study properly designed to address these questions would have ensured that the proportion of patients receiving treatment A would be identical (though not necessarily 1/2) within every category of interest. The post hoc separation of data into groups can destroy this balance and is generally to be avoided; it changes the most carefully designed experimental study into a possibly poorly designed observational one.

The observed reversals are characteristics of the particular probability distribution involved and, probabilistically, neither extraordinary nor surprising – such distributions are easily constructed and explained via the water container metaphor. From a scientific perspective, however, they can and should be avoided. Statistical design of experiments, including randomly allocating treatments to patients, provides important methodology which helps to ensure balance and hence to avoid such situations in scientific investigation (e.g. see Hand (1994) and Mackay and Oldford (2000) for further discussion on difficulties and on statistical method respectively). The investigators in this hypothetical study would be well advised not to rely on this study to draw conclusions about treatment effectiveness by either sex or age, but rather to explore these possibilities in future studies designed specifically to address such questions.



(a) Age within sex within treatment (b) Sex within treatment (c) No distinction

Figure 18: Double reversals: A is preferred to B for each grouping by age and sex; B is preferred to A for each grouping by sex; A is preferred to B absent any grouping. Applying the water container metaphor explains how this can happen. Moving through the figures from (a) to (c), left to right neighbouring pairs of containers are collapsed in one figure to produce the next figure.

9 The two envelope paradox

A wealthy patron presents you with two envelopes, indistinguishable from one another except that one is marked A and the other B. Inside one envelope is a cheque payable to you for some fixed amount, say x dollars; inside the other is another cheque to you for twice the amount, $2x$ dollars. Which envelope contains which cheque cannot be determined without opening both. You may choose either envelope and cash the cheque inside. Before you have opened the chosen envelope however, the patron will offer you the opportunity to change your mind and to take the other envelope instead – whatever your initial choice.

This simple problem has generated much public discussion relatively recently including, surprisingly, in professional newsletters and philosophy journals (e.g. Bickis 199x, Clark and Shackel, 2000). The discussion centres around the following paradoxical reasoning. Suppose you have chosen envelope A which contains some unknown amount – y dollars. The other envelope will have either half the amount (i.e. $y/2$ dollars) or twice the amount (i.e. $2y$ dollars). Each of these have equal probability of being the case. If you switch, half of the time you will get half as much and half the time twice as much. The expected amount you will receive if you switch is therefore $\frac{1}{2} \times \frac{y}{2} + \frac{1}{2} \times 2y = \frac{5}{4}y > y$. By switching, you will on average receive 25% more than you have in envelope A – you'd be crazy not to switch!

A moment's reflection tells you that something is wrong. Once you have B in hand, you can reason exactly as before and convince yourself that by switching from B to A the expected amount will again increase by 25% over whatever B contains. Having switched twice, you again have envelope A in hand, but now you also seemingly have something more – the mathematical assurance that the expected amount to be found there has just increased by over 56%! Simply by continually switching envelopes, back and forth between A and B, the expected amount can be increased beyond any level! Add a perpetual motion machine to keep the exchange going and the increase will never end. Something is wrong, but what?

To see, we first examine the correct solution, beginning with the outcome tree shown in Figure 19. The first branching corresponds to the initial placement of the money – A either has x dollars placed inside it or it has $2x$ dollars placed inside.The next branching corresponds to the decision

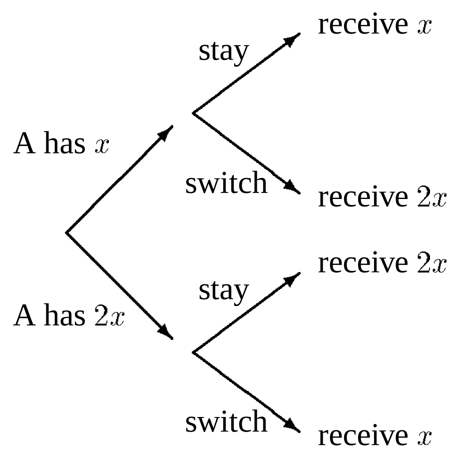
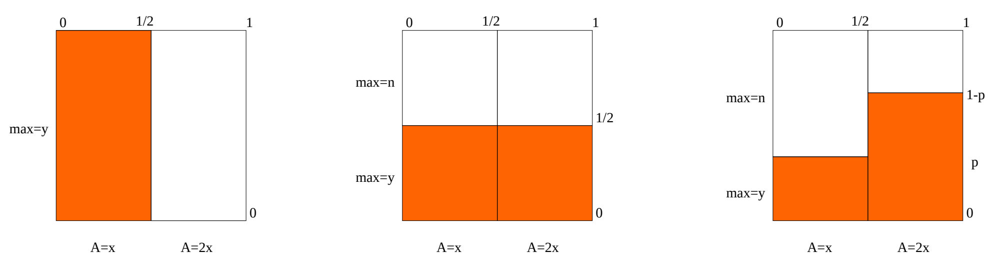


Figure 19: Envelope A has already been selected and contains either x dollars or $2x$ dollars. The question is whether to switch to envelope B, or to stay with envelope A.

either to stay with envelope A or to switch to envelope B. (For simplicity, the diagram assumes that you will always choose envelope A first but it could be easily modified to reflect the initial choice between envelopes by inserting another level of branching between the two levels which appear here; the argument would be unchanged.) At the leaves of the tree are the amounts that would be received by following those paths.

Different strategies, for deciding whether to stay with envelope A or to switch to B, can be examined by looking at the corresponding eikosograms. Figure 20 shows the eikosograms for



(a) Always switch (b) Switch with probability $\frac{1}{2}$ (c) Switch with probability p

Figure 20: Heights of shaded areas are the probability that the strategy has returned the maximum amount, $2x$, given that A $=x$, and given that A = $2x$. The vertical variable 'max' is 'y', for 'yes, the maximum is received' and 'n' for 'no the maximum is not received'. The total shaded area is the probability (whatever the vaue of A) that by switching the maximum, $2x$, is received. Note that in each of the three cases, this area is the same, namely 1/2.

three different strategies. A is just as likely to have x dollars as to have $2x$ dollars, so the strips corresponding to these two conditions have equal width in the eikosogram.

If, after first selecting envelope A, you always switch to B, then the probability that you will receive the maximum amount of $2x$ is one, if A had x inside, and zero, if A had $2x$ inside. This strategy is shown in the eikosogram of Figure 20(a). The total shaded area is the probability that you will receive $2x$ by this strategy and the unshaded area the probability that you will receive x. Each is 1/2 and the expected amount that will be received is $3x/2$, not $5x/4$. This is the same amount expected if you did not switch (shaded and unshaded regions of Figure 20(a) simply reverse roles). There is no paradox; you do just as well by staying with A as you do by switching to B.

The same holds for a variety of other strategies. If after selecting A, you opt for B with probability 1/2, then the eikosogram of Figure 20(b) results. The total shaded area again matches the total unshaded area and the expected amount received is unchanged at $3x/2$. Whatever the probability, p, of switching, as in Figure 20(c), the total shaded area and the total unshaded area are each 1/2 and the expected amount $3x/2$. No paradox arises.

Following the reasoning which led to the paradox, the outcome tree is rearranged as in Figure 21. Begin with A having some fixed, but unknown, amount y and then branch according to whether

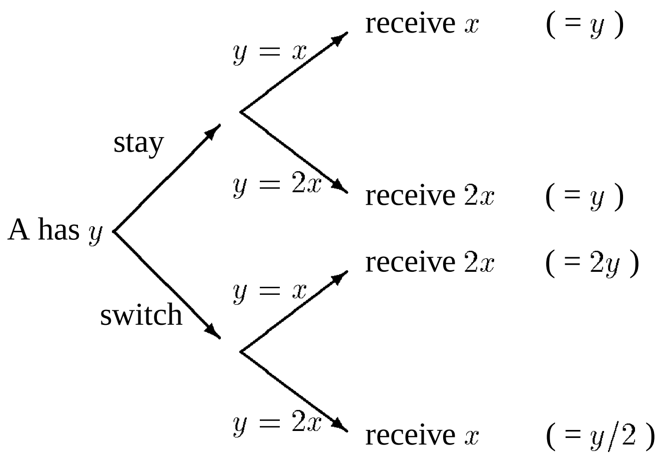


Figure 21: Outcome tree reordered to begin with A being selected initially and containing an unknown amount of money y. The first branching is now the decision to switch to B or stay with A. The second branching looks out into the future to see the amount to be received depending on the actual value of y.

the choice is to switch or not. If you switch, the lower branch is followed and the amount received is either twice ($2y$) the amount in A with or half that amount ($y/2$). Each of these having equal probability, the expected amount is calculated to be $5y/4$ or 25% more than the original amount y and the paradox has come about.

As indicated by the labelling of the outcome tree, the faulty reasoning is not so much about probability as it is about random variables. A calculated expected value of $5y/4$ assumes that the value of y is the same for each branch. However, as the tree labelling indicates, the *only* way which y can double is for it to have value x and the *only* way which y can halve is for it to have value $2x$. The *value* of y *is not the same* in the two cases. The correct expected amount received is therefore $\frac{1}{2} \times 2x + \frac{1}{2} \times x = \frac{3}{2}x$ as before.

While the outcome tree makes the story and its correct solution clear, some may prefer a strictly symbolic explanation which demands a more careful mathematical notation. As always, it is important to distinguish a random variable Y from its possible values y_i. We need also to recognize that amount received is a function of the value of y (and of the decision to switch or stay). If $r(Y = y_i)$ is the the amount received, when both Y has value y_i and you switch, then the expected amount received if you switch is denoted $E(r(Y)|switch)$ and determined as follows

$$\begin{aligned} E(r(Y)|switch) &= \sum_i r(y_i) \times Pr(Y = y_i|switch) \\ &= 2x \times Pr(Y = x|switch) + x \times Pr(Y = 2x|switch) \\ &= \frac{3x}{2} \end{aligned}$$

since Y can take only two possible values, x and $2x$. The second mistake of the paradoxical reasoning is to compare this amount to y, which is of course unknown. The correct comparison is to the expected amount received if you choose to stay with envelope A. This means following the top most paths of the outcome tree in Figure 21. The faulty calculation would produce $\frac{1}{2} \times y + \frac{1}{2}y = y$, seemingly justifying comparison to the unknown amount y. Denoting by $r^*(y_i)$ the return, when both $Y = y_i$ and you stay with A, the correct comparative calculation is

$$\begin{aligned} E(r^*(Y)|stay) &= \sum_i r^*(y_i) \times Pr(Y = y_i|stay) \\ &= x \times Pr(Y = x|stay) + 2x \times Pr(Y = 2x|stay) \\ &= \frac{3x}{2} \end{aligned}$$

which is the same as the expected amount received were you to switch. There is no advantage to switching or to staying and the paradox disappears.

This paradox is often embellished and made to seem more paradoxical by adding a probability distribution for the value x, the lower amount which the patron places in one of the envelopes. A remarkable amount of mathematics ensues including the possibility that x is infinite (e.g. Clark and Shackel, 2000). All of this is rendered immediately irrelevant by the simple fact that you are faced with the decision to switch or not *after* the patron has selected a value for x (which will never be infinite) and inserted the cheques in the envelopes. What value the patron might have chosen is irrelevant; some value has been chosen and the above reasoning applies. All probability calculations which follow are therefore *conditional* on x having been *fixed*. That the actual value of x is immaterial can be seen by changing the problem to one where instead of cheques, the envelopes contain a slip of paper with either 1 or 2 marked on it; the value of x is written on another slip of paper safely stored in a third envelope to be opened after the finally selected envelope A or B.

10 Monty Hall or the Prisoner's Dilemma

This is a now classic problem that has caused much debate in many fora. It is usually presented in one of two contexts: either as a once popular American television game show with host Monty Hall, or more sinisterly as a prisoner's dilemma where one of three prisoners is to be executed and the Monty Hall character is now cast as a knowledgeable guard. The debate centres around whether conditioning on seemingly irrelevant information actually changes any probabilities – it is a question of the independence of variables. As usually presented, there is not enough information to tell.

10.1 Monty Hall

The game show version begins with three closed doors behind one of which is a highly desirable prize such as a car. Behind each of the other two doors is a much less desirable prize, a goat. The game show host, Monty Hall, knows the location of the prizes; the contestant chooses a door expecting to receive the prize behind it. Before this door is opened however, Monty opens one of the other two doors to reveal a goat. The contestant, having seen this, is now presented the opportunity to change their mind: they can either stay with the door (still unopened) which they first selected, or they can switch to the other door which has not yet been opened. The point of contention is whether, having seen Monty's opened door reveal a goat, this information can now be used to advantage by the contestant to either remain with their original choice or to switch to the other unopened door.

That there should be no advantage is typically reasoned as follows. The contestant's original choice, say door C, has probability 1/3 of hiding the car. At this point, the contestant knows that one of the other two doors must hide a goat, the second might or might not. How could knowing that door B hides a goat be of any advantage? Surely, the car is still behind either door A or door C with equal probability? There is neither an advantage nor a disadvantage to switching from door C to door A.

To see that there could indeed be an advantage, requires a more careful examination. In probabilistic terms, we need to see whether

$$Pr\left(\text{Car is behind door C} \;\middle|\; \begin{array}{l}\text{Contestant selects door C and}\\ \text{Monty reveals a goat behind door B}\end{array}\right) \tag{1}$$

differs from 1/2. Of course, the information doesn't come to us in this order. The outcome tree of Figure 22 shows the possible outcomes of the game in time order left to right. Highlighted is the information that the contestant has chosen door C and Monty has revealed the goat behind door B. From this we can see that only two possible paths describe the known information and that only the bottom one of these yields the conclusion that the car is hidden behind door C. Note however that no probabilities have been assigned to these paths; jumping to the conclusion that the above probability must be 1/2 would be premature.

Figure 23 shows the eikosogram representing the known probabilities. As above, it assumes that the contestant has selected door C so that the entire eikosogram is conditional on this being the case. The conditioning variable takes a value A, B, or C when the car is truly behind the door of that letter; these probabilities are equal at 1/3. On the vertical axis is the probability that Monty reveals door B determined conditionally on the true location of the car.

If the car is behind door A and the contestant has already selected C then, as in the outcome tree, Monty has no choice but to reveal the goat behind door B - the conditional probability is one. Similarly, if the car is behind door B then it cannot be revealed and the conditional probability must zero – the goat behind door A must be revealed. However, when the true location is door C, the door already selected by the contestant, then Monty has some choice – he can choose to open door B with any probability he likes between zero and one. If this probability is p, say, then door A will be opened with probability $1 - p$. In terms of the outcome tree, it matters what probability value will be attached to the last leaf of the lowest branch.

The desired conditional probability is now easily determined from the eikosogram. It is the

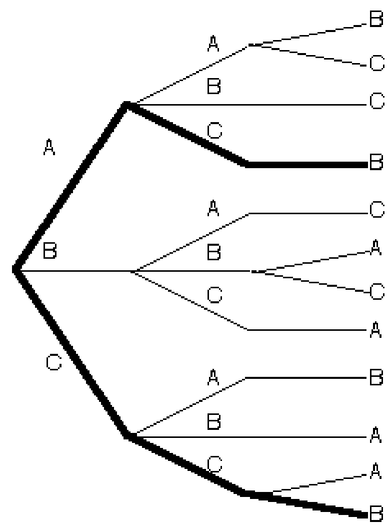


Figure 22: Outcome tree for the Monty Hall problem. From the left: first the car is assigned to one of doors A, B, or C; second, the contestant chooses one of the doors; third, Monty reveals the goat behind one of the other doors. Shown with thick lines is the event that the contestant has chosen door C and Monty has revealed the goat behind door B; the car must be behind either door A or door C.

ratio of the rightmost shaded area to the total shaded area, or

$$Pr\left(\text{Car is behind door C}\;\middle|\;\begin{array}{l}\text{Contestant selects door C and}\\ \text{Monty reveals a goat behind door B}\end{array}\right) = \frac{p}{1+p} \tag{2}$$

which has value 1/2 if, and only if, $p = 1$ – that is, if and only if Monty never opens door A but always opens door B when the contestant's door C hides the car. In this case, from the contestant's perspective it doesn't matter whether they switch from door C to door A or not – the probability of getting the car is the same.

Alternatively, if in these circumstances Monty always opens door B and never door A, then $p = 0$ and the conditional probability is zero. The car must be behind door A and the contestant would be wise to switch.

These are the two extremes. If, as many assume to be the case, Monty chooses to open door A and door B with equal probability. i.e. $p = 1/2$, then the probability that the contestant's selection hides the car is 1/3 and that door A hides the car is 2/3 – the contestant would again be wise to switch.

Not knowing Monty's rule, and so not knowing the value of p, prevents us from declaring that there is, or is not, an advantage to switching. Nevertheless, the ambiguity can be resolved to the satisfaction of the contestant. Having p take any value between its two extremes amounts to setting the height of the rightmost shaded bar In Figure 23. As is plain from the picture, no value of p will make the relative area of the rightmost shaded region to the total shaded region greater than 1/2. That is, the conditional probability is never greater than 1/2 and could be much less than 1/2. Switching from door C to door A, the contestant can never reduce the probability of getting the car and just might increase it. In the absence of any knowledge of p, it makes sense to switch.

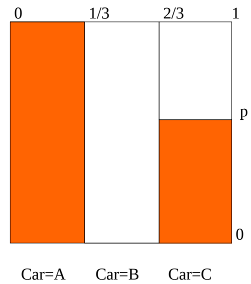


Figure 23: The Monty Hall Problem: The diagram assumes the contestant has already chosen door C. Monty will then open one of the other two doors to reveal the goat. The conditioning variable is the true location of the car. Shaded areas indicate the probability that Monty opens door B, unshaded areas the probability that Monty opens door A.

10.2 The prisoner's dilemma

The prisoner's dilemma[3] is identical to the Monty Hall problem in its outcome structure. The context now is that three people, whom we will call Al, Bob, and Carl to emphasise the connection with Monty Hall's doors A, B, and C, are being held prisoner. In a gross violation of humanitarian principles, one has been 'randomly selected' (i.e. with equal probability) to be executed in the morning. The name of this person will be revealed in the morning and each prisoner, unable to communicate with the others, is left to worry through the night about what might await them come dawn.[4]

There is a rather sadistic guard (coincidentally named Monty) who, knowing the identity of the prisoner to be executed, informs Carl that Bob is not to be executed. The question is, should Carl's concern about his fate have changed in light of this information?

The outcome tree of Figure 22 serves to describe this situation as well and application of the same analysis leads to the corresponding eikosogram for Carl's fate as shown in Figure 24 As can be seen it is identical in structure to that of Figure 23 for the Monty Hall problem. However, instead of considering whether to switch, Carl's sole concern is to determine the conditional probability that he will be executed given this additional information.

As with the Monty Hall problem, this conditional probability is the ratio of the shaded area on the right of Figure 24 to the entire shaded area. From the eikosogram this is

$$Pr\left(\text{Carl is to be executed}\,\middle|\,\text{Guard says Bob will be spared}\right) = \frac{p}{1+p} \tag{3}$$

just as before.

[3]There is a classic problem in game theory which goes by the same name and should not be confused with this dilemma. The game theoretic 'prisoner's dilemma' has two separate prisoners (who were partners in a crime) each trying to decide whether to confess or not. If both confess, they will both serve time; if only one confesses, implicating the other, the confessor will have a much reduced sentence; if neither confess then neither serves time. No one is killed.

[4]Less disturbing, although hopefully no more realistic, versions of this problem exist whereby, rather than execution, release of one or two of the prisoners occurs.

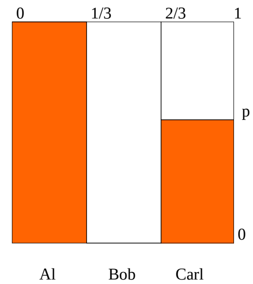


Figure 24: The Prisoner's Dilemma: The diagram assumes the prisoner is Carl. The guard will then name one of the other two prisoners as one who who will go not be executed. The conditioning variable is the identity of the prisoner to be executed. Shaded areas indicate the probability that the guard names Bob, unshaded areas the probability that the guard names Al.

The difficulty is that once again we do not know how the guard chooses to name Bob over Al when Carl is the one to be executed. If the guard names Bob or Al with equal probability, then the conditional probability is 1/3 and Carl has no more worry than he had before the guard named Bob. If, however, Bob is always named in such circumstances (e.g. perhaps the guard really knows only that Bob is to be spared and not the identity of the one to be executed) then the conditional probability is 1/2 and Carl has more to worry about. The conditional probability can range anywhere from zero to 1/2, depending on the value of p. In the absence of further information about p Carl has no more reason for increased concern than he has for increased comfort. Carl might take some solace from the fact the favourable interval $[0, 1/3]$ is larger than the unfavourable interval $[1/3, 1/2]$, but this would only be reasonable if Carl had yet further information, namely that the guard was just as likely to choose a p between 0 and 1/2 as between 1/2 and 1.

A paradox is sometimes asserted by suggesting that any choice other than $p = 1/2$ seems to mean that the guard, merely by whispering "Bob" to Carl, or to himself for that matter, can change Carl's fate by increasing or decreasing the probability of his execution. This being absurd, it is argued (e.g. Isaac, 1995, p. 26) that to render the problem realistic and to remove the paradox a further condition must be assumed – we must assume that the prisoner's fate is independent of the guard's statement and so, that the conditional probability of execution is the same as the unconditional probability, which is 1/3. This in turn forces $p = 1/2$ and resolves the 'paradox' by appeal to intuition. Because the apparent paradox is unacceptable, conditions have been assumed which necessarily preclude it.

But of course there is no paradox to resolve. The above argument has confused *causal* independence with *probabilistic* independence. The guard's actions no more change Carl's fate than does Monty Hall's revelation cause the goats and car to change positions; these are established at the outset. Moreover, the guard could whisper "Carl" just as easily as "Bob". Neither would affect Carl's fate even though, in the former case, it would become clear to Carl that he is not the one to be executed – no doubt remains. What the guard's words have changed is the information available to Carl. Carl, knowing the structure of the problem and some of the probabilities involved, can revise the probability that someone in his state is executed given the information available. As information accumulates, the conditions of the conditional probability adjust accordingly and the value can change.[5] Setting $p = 1/2$ a priori is unjustified; the conditional probabilities are as they are without any implication that the guard has affected Carl's fate, only that he has affected Carl's information.

10.2.1 The difference and a revised prisoner's dilemma

There is one difference in structure between the prisoner's dilemma and the Monty Hall problem which needs pointing out. In the Monty Hall problem the contestant selects one of the three doors without knowing which contains the car; in the prisoner's dilemma the guard chooses to speak with the prisoner knowing full well who will be executed. This choice could also be made part of the problem.

The difference occurs at the second branching in the outcome tree, shown again In Figure 25 but now with the probability p and two new probabilities a and c added to represent the conditional

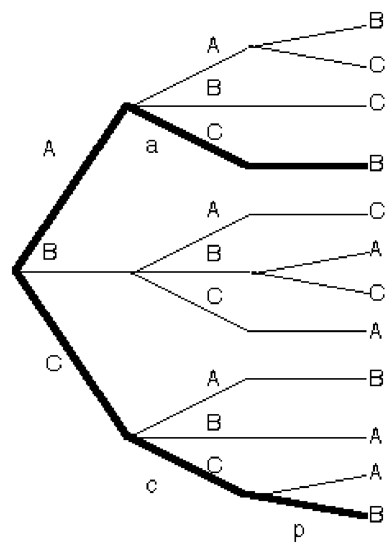


Figure 25: Outcome tree for a revised prisoner's dilemma. Probabilities a and c are the conditional probabilities that the guard provides information to Carl (C) when Al or Carl, respectively, is the one to be executed.

probability that the guard gives information to Carl, respectively when Al is to be executed and when Carl is to be executed.[6] In the Monty Hall problem these would be the conditional proba-

[5]To give a more serious example, consider a family history, H, of occurence of a disease and a genetic marker, M, which is known to increase the incidence of the disease when present. We can talk about $Pr(H|M)$ via the mechanics of inheritance. In diagnosis, however, it is more interesting to determine $Pr(M|H)$, the probability that the family has the genetic marker given its history of disease. The fact that the marker causes the disease and not the other way around does not preclude the determination of this probability of interest. Moreover, it would be absurd (and possibly dangerous) to use the causal direction to assert that $Pr(M|H)$ must be the same as the unconditional $Pr(M)$. In the same way, the probability of Carl's execution given the information from the guard need not be the same as his unconditional probability; were it necessarily so, no learning would be possible.

[6]Another probability, b, could be given when Bob is to be executed but no use will be made of that here.

bilities that the contestant chooses door C given that the car is behind doors A and C respectively. The difference between the two problems then is that for the contestant, having no information on the location of the car, a must equal c, whereas for the guard they could be entirely different. In the previous analysis it was assumed that they were both 1.[7]

If the guard has the probabilities a and c as shown, then the eikosogram of Figure 24 for the prisoner Carl now becomes that of Figure 26. The strip representing the condition that Al is to be

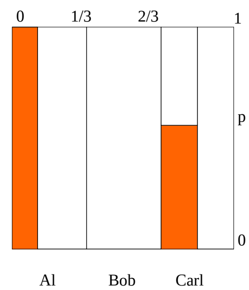


Figure 26: The Prisoner's Dilemma: The diagram is as in Figure 24 except that the strip for Al and that for Carl from Figure 24 are here each split according to the probabilities a and c respectively. If Al is to be executed then two thin strips at the left apply – the first has width $a/3$ representing the probability that the guard talks to Carl and Al is to be executed; the second the probability $(1 - a)/3$ that the guard does not talk to Carl and Al is to be executed. Similarly the two thin strips associated with Carl have width $c/3$ and $(1 - c)/3$ left to right resepecively.

executed is now split in two – the left of width $a/3$, the right of width $(1 - a)/3$ – according to whether the guard talks to Carl or not. The height of the shaded bar still represents the conditional probability that the guard names "Bob" as being spared; the heights of the unshaded portions are conditional probabilities that "Bob" is not named (meaning either Al or no one was named depending on the condition). The strip when Bob is to be executed could have been split as well according to the conditional probability that the guard speaks to Carl but since the guard will never name Bob inunder these conditions, the splitting is irrelevant.

The conditional probability of interest is again the ratio of the right most shaded area to the total shaded area, now seen to be

$$Pr\left(\text{Carl is to be executed} \,\middle|\, \text{Guard tells Carl that Bob will be spared}\right) = \frac{cp}{a + cp} \quad . \tag{4}$$

The analysis reverts to what we have seen before whenever $a = c$ (Monty Hall has $a = c \neq 0$; the prisoner's dilemma $a = c = 1$). That a and c cannot be simultaneously zero is implied by the fact that the guard has spoken to Carl.

As is clear from either Figure 26 or Equation (4), without further restriction on the values of a, c and p, this conditional probability can be any value between 0 and 1 inclusive. Given only the

[7]For the prisoner's dilemma the probabilities need not sum to one across the three second-level branches A, B, C because the guard can speak to each prisoner. For the Monty Hall problem, these probabilities must sum to 1 since only one of the three doors can be selected.

probability structure without restriction on its values, Carl can draw no conclusions whatever from the guard's information and would be better to ignore it.

Suppose now that the guard, being somewhat sadistic, informs Carl that he, the guard, is twice as likely to give this kind of information (i.e. Bob being spared) to the prisoner who is condemned to die than to a prisoner who will be spared, i.e. $c = 2a$. From this Carl can determine that the conditional probability is $2p/(1 + 2p)$. Now if the guard says that when Carl is to be executed he will name Al or Bob with equal probability, Carl's conditional probability has become 1/2; he has learned something from the guard's information. Had the guard been ten times as likely, i.e. $c = 10a$, this conditional probability would have become 5/6.

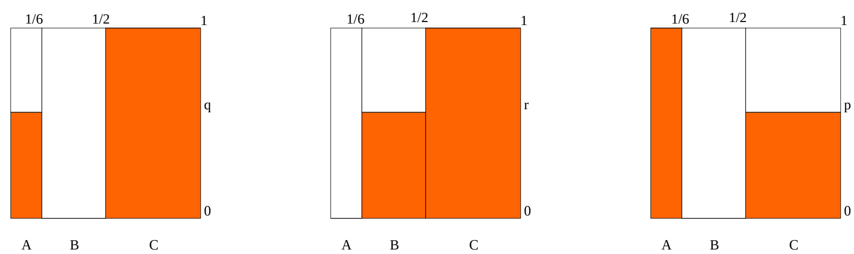
Ignorance may not be bliss but it can be a good deal less worrying.

10.3 A twist on Monty

Although the differing probabilities a and c, seen in the revised prisoner's dilemma, cannot be used in the Monty Hall problem, nothing prevents the game show from assigning the car to the three doors with different probabilities. To be concrete, suppose the car is assigned to door A with probability 1/6, door B with probability 1/3 and door C with probability 1/2. To be fair, the contestant will also know these probabilities. Everything else is as before.

Knowing that the car is placed behind door C fully half the time, it is easy to imagine that a large percentage of people would select door C at once. But is this the wisest strategy knowing that, whatever the selection, Monty will always offer the contestant the opportunity to switch?

Figure 27 shows the three eikosograms corresponding to the contestant having first selected



(a) Door A selected (b) Door B selected (c) Door C selected

Figure 27: Shaded heights give conditional probability (given the car is behind the doors as named on the horizontal axes) of Monty revealing a goat behind (a) Door B, (b) Door A, and (c) Door B.

each of the doors A, B, or C. In Figure 27(b), door B cannot be revealed, since it is the door selected, so the conditional probabilities shown by the heights of the shaded bars will be that Monty reveals a goat behind door A; Figures 27(a) and (b) will have Monty revealing the goat behind door B as before. To be general, three possibly different probabilities q, r and p will represent the conditional probabilities for revealing the goat when Monty has a choice between doors.

The following three probabilities are determined as before by the ratios of the relevant shaded areas:

$$Pr\left(\text{Car is behind door A}\;\middle|\;\begin{array}{l}\text{Contestant selects door A}\\ \text{and Monty reveals a goat}\\ \text{behind door B}\end{array}\right) = \frac{q/6}{1/2+q/6} = \frac{q}{3+q}$$

$$Pr\left(\text{Car is behind door B}\;\middle|\;\begin{array}{l}\text{Contestant selects door B}\\ \text{and Monty reveals a goat}\\ \text{behind door A}\end{array}\right) = \frac{r/3}{1/2+r/3} = \frac{2r}{3+2r}$$

$$Pr\left(\text{Car is behind door C}\;\middle|\;\begin{array}{l}\text{Contestant selects door C}\\ \text{and Monty reveals a goat}\\ \text{behind door B}\end{array}\right) = \frac{p/2}{1/6+p/2} = \frac{3p}{1+3p}\quad .$$

It also matters which door Monty opens. The following three probabilities are calculated by forming the relevant ratios of the *un*shaded areas of Figure 27(a), (b) and (c):

$$Pr\left(\text{Car is behind door A}\;\middle|\;\begin{array}{l}\text{Contestant selects door A}\\ \text{and Monty reveals a goat}\\ \text{behind door C}\end{array}\right) = \frac{(1-q)/6}{1/3+(1-q)/6} = \frac{1-q}{3-q}$$

$$Pr\left(\text{Car is behind door B}\;\middle|\;\begin{array}{l}\text{Contestant selects door B}\\ \text{and Monty reveals a goat}\\ \text{behind door C}\end{array}\right) = \frac{(1-r)/3}{1/6+(1-r)/3} = \frac{2-2r}{3-2r}$$

$$Pr\left(\text{Car is behind door C}\;\middle|\;\begin{array}{l}\text{Contestant selects door C}\\ \text{and Monty reveals a goat}\\ \text{behind door A}\end{array}\right) = \frac{(1-p)/2}{1/3+(1-p)/2} = \frac{3-3p}{5-3p}\quad .$$

Given values of p, q and r the contestant can develop a strategy for the game. If no values are available to the contestant, then as before the strategy will need to assume all values are possible.

In Table 10.3 the range of possible values for these probabilities is given in the third column. If

Door Selected	Door Showing Goat	Probability of getting the car: Range, Stay	Range, Switch	$p = q = r = 1/2$, Stay	$p = q = r = 1/2$, Switch
A	B	[0, 1/4]	[3/4, 1]	1/7	6/7
A	C	[0, 1/3]	[2/3, 1]	1/5	4/5
B	A	[0, 2/5]	[3/5, 1]	1/4	3/4
B	C	[0, 2/3]	[1/3, 1]	1/3	2/3
C	A	[0, 3/5]	[2/5, 1]	3/7	4/7
C	B	[0, 3/4]	[1/4, 1]	3/5	2/5

Table 5: Some conditional probability values for the twisted Monty.

the contestant does not switch but stays with the door first selected, then these become the ranges for the conditional probabilities of receiving the car. Similarly, the fourth column provides the ranges for the probabilities of winning the car under each condition when the contestant switches.

The last two columns are the corresponding values when Monty has decided to flip a coin (i.e. $p = q = r = 1/2$) to decide which door to open whenever he has a choice.

Consider the last two columns first. In every case but one, the contestant is better off switching as that increases the probability of getting the car. Moreover it matters which door is first selected. As the last two columns show, nowhere is the probability of their winning a car better than when they first select door A and then switch, whichever door Monty shows. The lowest probability in this case is 4/5 or 80% and is larger than any probability asssociated with selecting either door B or C at the start. The next best choice is door B, with door C last (even following a mixed strategy of switching when Monty opens A and not when he opens B). The ordering is the opposite of what one might first suppose.

If the values of p, q, anr r are unknown then a similar logic applies where the contestant chooses to maximize the minimum probability of their getting the car. Column four of Table 10.3 shows that again the best choice is to select door A first and then switch – the probability of getting the car is at 2/3 and possibly higher. By contrast the worst choice would be door C, where the largest minimum probability is 1/4.

So whether the contestant knows that Monty chooses which door to open "at random" when he has a choice or uses some other unknown probabilities, the recommended strategy is not to choose door C and hold that choice, but to choose door A and switch.

11 Discussion

Acknowledgements

References

Baker, S.G. and B.S. Kramer (2001). "Good for Women, Good for Men, Bad for People: Simpson's Paradox an the Importance of Sex-Specific Analysis in Observational Studies", *Journal of Women's Health & Gender-Based Medicine, 10(9)*, pp. 867-872.

Bell, D.E., Raiffa, H. and A. Tversky, editors (1988). *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, Cambridge University Press, Cambridge.

Bar-Hillel, M. (1980) "The base-rate fallacy in probability judgements", *Acta Psychologica*, *44*, pp. 211-233.

Bickis, M. (199x) "Ali-Baba???", *Liaison*, *??*, pp. ??-??.

Cadarette, S.M., Jaglal, S.B., Kreiger, N., McIsaac, W.J., Darlington, G.A. and J.V. Tu (2000). "Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry", *Canadian Medical Association Journal*, *162(9)*, pp. 1289-94.

Cherry, W.H. and R.W. Oldford (2003) "Picturing Probability: the poverty of Venn diagrams, the richness of eikosograms", *submitted for publication*.

Clark, M. and N. Shackel (2000). “The Two-envelope Paradox”, *Mind, 169(435)*, pp. 415-442.

Hand, D.J. (1994). “Deconstructing Statstical Questions (with discussion)”, *Journal of the Royal Statistical Society, Series A (Statistics in Society), 157(3)*, pp. 317-356.

Isaac, R. (1995). *The Pleasures of Probability*, Springer-Verlag, New York.

Kahneman, D. and A. Tversky (1972). “On prediction and judgement”, *ORI Research Monograph, 12(4)*.

Kahneman, D., Slovic, P. and A. Tversky, editors (1982). *Judgement under uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge.

Mackay, R.J. and R.W. Oldford (2000). “Scientific Method, Statistical Method and the Speed of Light”, *Statistical Science, 15(3)*, pp. 254-278.

McNeil, B.J., Pauker, S.G., Sox, H.C. Jr., and A. Tversky (1982). “On the Elicitation of Preferences for Alternative Therapies”, *New England Journal of Medicine*, 306, pp. 1259-1262.

McNeil, B.J., Pauker, S.G. and A. Tversky (1988). “On the Framing of Medical Decisions”, Chapter 20 of Bell et al, 1988, pp. 562-568.

Oldford, R.W. (2003) “Understanding probabilistic independence and its modelling via eikosograms and graphs”, *submitted for publication*.
Osteoporosis Society of Canada (2003). “About Osteoporosis”, *www.osteoporosis,ca*, March 2003.

Tversky , A. and D. Kahneman (1982). “Evidential impact of base rates”, Chapter 10 of Kahneman, et al (1982), pp. 153-160.

Tversky , A. and D. Kahneman (1988). “Rational Choice and the Framing of Decisions”, Chapter 9 of Bell et al (1988), pp. 167-192.