Understanding Conditional Probability

KEYWORDS:
Teaching;
Reasoning;
Venn diagram;
Tree diagram

Stephen Tomlinson
University of Alabama, USA

Robert Quinn
University of Nevada, USA

Summary
This article offers a new approach to teaching the difficult concept of conditional probability.

◆INTRODUCTION◆

THE fundamental aim of education is the development of understanding: the mastery of concepts and procedures that permit the intelligent solution of problems in new and novel situations. In mathematics this goal is particularly challenging, for students are often adept at memorising the rules and formulas necessary to solve well-defined test and textbook problems, without ever coming to terms with the meaning or logic of the arguments they employ. As research shows, outside of the classroom even the most familiar operations can be ignored in favour of intuitive judgements (Gardner, 1991; Piattelli-Palmarini, 1994). But while commonsense may provide serviceable solutions to many problems, experience also reveals that our *natural* theories can lead us into serious errors. To improve their intellectual skills, students must learn to recognise the limits of their spontaneous problem-solving strategies and develop *secondary* intuitions for the more precise and powerful reasoning tools developed by mathematicians (Fischbein, 1987).

◆PROBABILITY AND ◆ JUDGEMENT

One of the most difficult areas in which to achieve these goals is probability theory. On one hand, the laws of probability contain many abstract expressions, complex terms, and nested relationships that pupils find hard to understand. On the other hand, psychologists, building upon the pioneering work of Amos Tversky and Daniel Kahneman, have shown that all human beings approach experience with expectations and stereotypes that exert a strong and oftentimes distorting grip over judgement (Tversky and Kahneman, 1974). Consider, for example, how the mind imposes a script on the following problem: If Linda is a thirty-one year old single woman who is outspoken on social issues such as disarmament and equal rights, which of the following statements is more likely to be true?

P: Linda is a bank teller.
Q: Linda is a bank teller and active in the feminist movement.

According to Kahneman and Tversky more than 80% of those questioned – including many schooled in statistics – chose *Q*, even though the set of bank tellers who are feminists is included within the set of bank tellers, and is thus a smaller proportion of the population (i.e. since $Q \subseteq P$ then $p(Q) \leq p(P)$) (Kahneman, Slovic and Tversky, 1982). We posed the same question to 18 undergraduate mathematics education majors and found that only two judged *P* the more probable event!

The distorting influence of our natural reasoning skills can also be seen in the way people tend to systematically ignore relevant knowledge in the decision-making processes. For example, to test how subjects integrated “base-rates” into their judgements, Tversky and Kahneman presented two groups with 100 character profiles. Group A was told that their dossiers consisted of 70 lawyers and 30 engineers, while group B was informed there were 70 engineers and 30 lawyers. Yet, despite these different populations, each group divided their reports in roughly the same proportion – even though group A should have identified more than twice as may lawyers as group B! Interestingly, when given a stack of neutral personality sketches, such as:

> "Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues."

both groups split the lawyers and engineers fifty-fifty not seventy-thirty or thirty-seventy as would be expected (Tversky and Kahneman, 1974).

According to Howard Gardner the cognitive heuristics that naturally guide reasoning in such cases are rather like optical illusions - pathways in thought that lead to conclusions which, upon reflection, can be recognised as erroneous. And yet, despite their intuitive necessity, Gardner believes such judgements can be overcome through meta-cognition, the conscious use of mathematical tools. Just as we can employ the rules of logic to construct valid arguments, so, he argues, a knowledge of probability can help us formulate sound inferences. The problem is that the ability to employ such secondary reasoning is not fostered by traditional teaching practices resting upon the memorisation of formulas. In recent years educators interested in stochastics have responded to this challenge by constructing numerous pedagogical strategies that help students recognise the limits of their subjective theories and learn how to explore probabilistic relationships objectively. For instance, many papers included in the proceedings of the first and second International Conferences on Teaching Statistics (ICOTS) stress the need for pupils to conduct experiments, play games, employ computer simulations, and use graphical representations in order to gain an empirically grounded measure of events (Grey, Holmes, Barnett and Constable, 1983; Davidson and Swift, 1988). Such concrete activities are necessary to generate what might be called a "Socratic encounter" with the laws of probability; they provide a level of personal investment in the resolution of problematic situations that permits serious questioning of judgement and an openness to the consideration of mathematical laws. However, if students are to understand and employ normative methods of problemsolving to enhance their decision-making these practical explorations must be augmented with the theoretical modelling of events, for, whether *frequentist* or *classical*, it is only through an intuition of mathematical arguments that intelligent thought and rational judgement can be promoted.

Yet, as research reveals, students find the notion of independent events and the laws of conditional probability extremely difficult to understand (Shaughnessy, 1993). In the following discussion we offer a new approach to these concepts that can help students develop a working understanding of probability laws. By representing events pictorially, we show how the Venn diagram and the Tree diagram can be integrated into a single problem-solving instrument that provides a graphical mechanism for constructing conditional probabilities and answering Bayesian questions without having students memorise complex formulas. Since conditional relationships and Bayes' Theorem lie at the heart of the scientific method, and indeed permeate everyday experience, the development of such reasoning skills is vital for informed decision-making in many situations.

◆VENN DIAGRAMS◆

Consider the following three events based upon a single roll of a fair die:

X: An even number
Y: A number greater than two
Z: A prime number

While Venn diagrams are helpful in sorting out which outcomes in an experiment are common to different events, they do not clearly reveal the underlying relationships of dependence and independence that are essential to the solution of conditional probability problems. Events may be logically or contingently connected to one another, in such a way that the occurrence of one increases or decreases the probability of the other. Thus, while the chance of rain is *independent* of our wishes, the number of people carrying umbrellas to work *depends* upon the weather forecast. In the case of the events X, Y and Z, however, such relationships are not at all obvious. Does knowledge that a prime has been rolled increase, decrease, or not affect the probability of the outcome being an even number? Further, given such conditions, how can the probability of getting both a prime and an even number be determined? As the story about Linda demonstrates, questions like these, which are central to many probability problems, can only be answered when the implicit relationships between events are fully understood.

As Fig. 1 illustrates, it is possible to find the probability of combined events like *"A* and *B"*, hereafter *p(AB)*, by exhaustively arranging all the simple events within the appropriate regions of the Venn diagram. But, importantly, this value does not

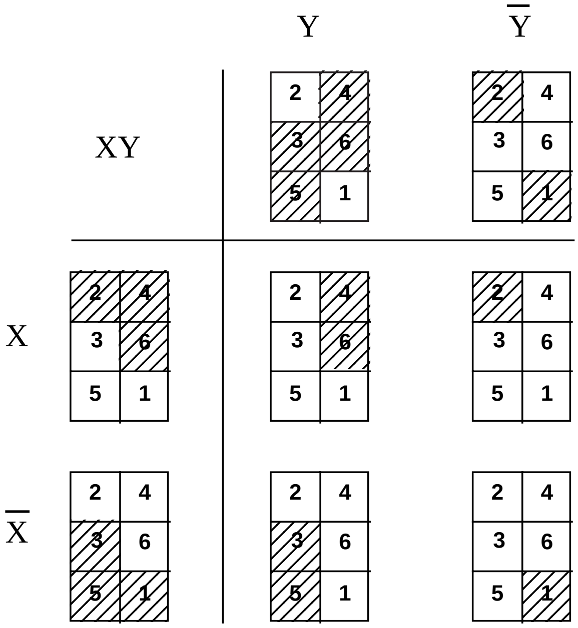
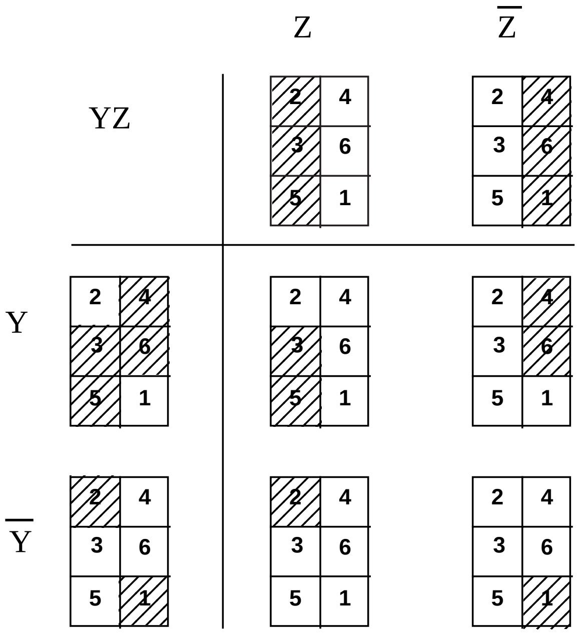


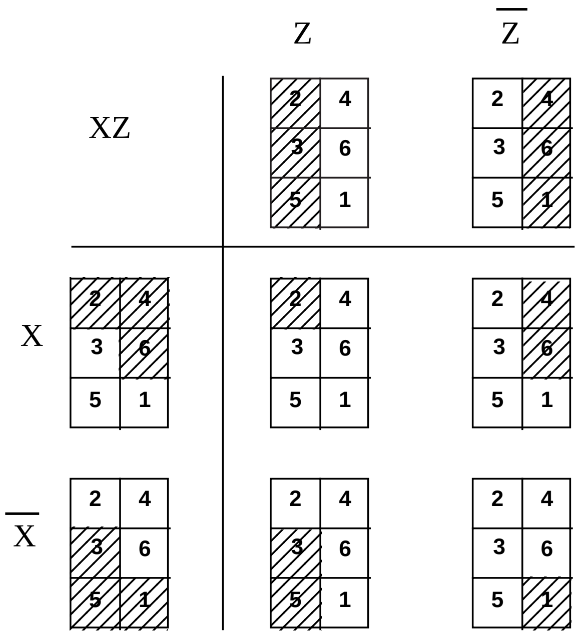
Figure 1 Intersection Tables for sets X, Y and Z

X: an even number. Y: a number greater than 2, Z: a prime number

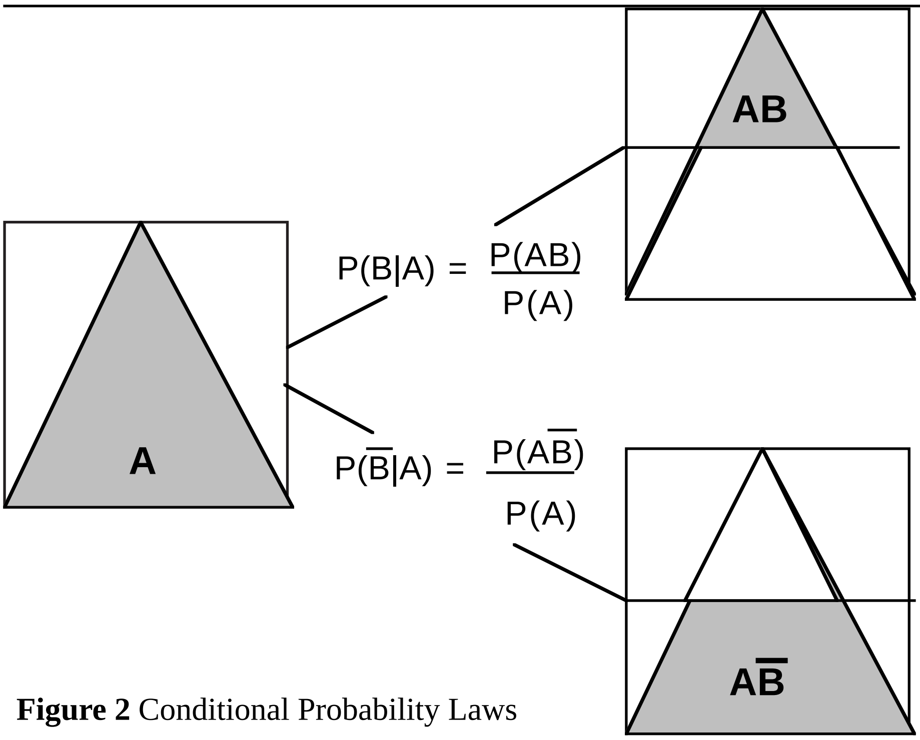
XY	p(Y)=2/3	p($\bar{Y}$) = 1/3
p(X) = 1/2	p(XY) = 1/3	p(X$\bar{Y}$) = 1/6
p($\bar{X}$) = 1/2	p(XY) = 1/3	p($\bar{X}$ $\bar{Y}$) = 1/6



YZ	p(Z)=1/2	p($\bar{Z}$) = 1/2
p(Y) = 2/3	p(YZ) = 1/3	p(Y $\bar{Z}$) = 1/3
p($\bar{Y}$) = 1/3	p($\bar{Y}$ Y) = 1/6	p($\bar{Y}$ $\bar{Z}$) = 1/6



XZ	p(Z)=1/2	p($\bar{Z}$) = 1/2
p(X) = 1/2	p(XZ) = 1/6	p(X $\bar{Z}$) = 1/3
p($\bar{X}$) = 1/2	p($\bar{X}$ Y) = 1/3	p($\bar{X}$ $\bar{Z}$) = 1/6



If set A is represented by the triangle and set B by the upper rectangle:

p(A) =1/2

p(AB) = 1/8

p(B) = 1/2

p(N|A) = p(AB)/p(A)

= (1/8)/1/2) = 1/4

Figure 2 Conditional Probability Laws

always equal the product *p(A)p(B),* as the symbols *"AB"* or the term "and" might suggest. For example, while the "intersection tables" for *XY* and 17 show that p(XY) = p(X)p(Y) and p(YZ) = p(Y)p(Z), the table for XZ reveals that p(XZ) ≠ p(X)p(Z). This is because X and *Z* are dependent events influencing one another's probability while events X and Y, and events Y and *Z* are both independent.

As teachers will recognise, one of the most common assumptions students make is that the probability of a combined event can be calculated by multiplying its constituent probabilities. While this empirical "product rule" holds for independent events, such as Y and *Z,* students must come to appreciate how this formula breaks down with dependent relationships. Having pupils postulate and then compute the values p(X), p(XY), p(XY), and so on, for every entry in these tables will lead them to face this fact squarely and, as we have argued, provide a powerful stimulus for motivating the mathematical exploration of these events. Clearly, having students memorise and apply the formula *P(AB) = p(A)p(B|A)* in textbook exercises will not ensure their grasp of probability laws. Because the expression *p(B|A),* the probability of B given A, is a comparison of the simple events in set A and the simple events in its subset *AB,* the underlying relationships between events must be understood through the subset relation. Indeed, *all probabilities are conditional,* and are computed from the same basic definition:

p(Event) = p(Solution set)/p(Outcome Set)

Even event *Z,* rolling a prime number, depends upon the die being fair. Thus,

p(Z) = p(Z|Fair die) = p(Z|Universal Set) =

$$\frac{\text{the sum of the simple events in set } Z}{\text{the sum of the probabilities in set U}}$$

$$= \frac{3/6}{6/6} = 1/2$$

This argument can be presented pictorially by making the area of each region within the Venn diagram correspond to the probability of the event it represents. Thus, the universal set has an area of 1, and, in the example above, set *Z* has an area of 1/2.

$$p(Z) = \frac{p(AB)}{p(A)} = \frac{\text{Area of region } AB}{\text{Area of region } A}$$

In general, for any two events A and B, if p(A) ≠ 0

$$p(Z) = \frac{p(AB)}{p(A)} = \frac{\text{Area of region } AB}{\text{Area of region } A}$$

Thus, from figure 1

$$p(Z \mid X) = \frac{p(ZX)}{p(X)} = \frac{1/6}{1/2} = 1/3$$

But how are values like p(ZX) determined without exhaustively enumerating all the simple events in a Venn diagram? Here we must turn to the second of our mathematical tools, the Tree diagram.

◆TREE DIAGRAMS◆

The Tree diagram is most commonly used to determine the probabilities of events in a multistage experiment. For example, the probability of getting two heads when a coin is tossed three times, or the probability of randomly selecting two green marbles from ajar that contains one blue, three green, and two white ones. In such cases, the dependence or independence of events is clearly defined by the causal sequence of the experiment. It is easy to see that the result of tossing one coin does not influence how the next one will land, but that removing a blue marble from the jar does alter the probability of future selections. The difficulty comes with events such as X, *Y* and *Z,* where conditional relationships must be determined by looking at the nest of *logical* connections defined by the structure of their subsets. We therefore propose that Venn diagrams be integrated with the Tree diagram by defining the various branches to represent an ordered proper subset relation on exclusive events. (When read in the opposite direction, the Venn diagrams of each stage thus become the union of sets at the previous level). The subsets of A are AB and A $\overline{B}$, while, conversely, $AB \cup A\overline{B} = A$.The basic atom of this scheme, defining the relationships *between p(A), p(AB)* and *p(B|A)* is shown in Fig. 2. Further, as this diagram illustrates, by partitioning the unit square into "event regions", area can be used to represent the probability values of all simple and combined events.

◆BAYES' THEOREM ◆

An important class of questions turns upon the computation of conditional probabilities that are *implicit* in an experiment or a set of data. For example, one might ask "What is *p*(Green on first draw| Blue on second)?" Like asking for *p(A|B)* in Fig. 2, this problem inverts the direction of the tree diagram; if the branches lead from *A* to *B*, how can a probability be computed that depends on a knowledge of B prior to A? As Falk shows, when faced with such "diagnostic" questions (so named because Bayesian relationships often arise in medical testing), most students have little or no idea of how non-causal connections can be computed - some even believe that the problems themselves are nonsensical (Falk, 1994). The traditional approach to such questions is to employ Bayes' Theorem, which in its simplest form states

$$p(A \mid B) = \frac{p(A)p(B \mid A)}{p(A)p(B \mid A) + p(\overline{A})p(B \mid \overline{A})}$$

However useful this formula may be, it clearly does not provide the student with an intuition of the reasoning process necessary to solve such embedded problems. Yet the expression *p(B|A)* depends upon a subset relation similar to those in the previous conditional probability problems. That is,

$$p(A \mid B) = \frac{p(AB)}{p(B)} = \frac{\text{Area of region AB}}{\text{Area of Region B}}$$

The difficulty, of course, is that there is no Venn diagram for set *B* in a tree diagram that starts with event *A.* If there were, *p(A|B)* could be read off as easily as *p(B|A).* But, as we have argued, since the tree diagram is composed of the mutually exclusive events whose union is the Universal set, it does contains the building blocks out of which new Venn diagrams can be constructed. In short, set B - and set $\overline{B}$ - can be produced very simply by extending the tree using the operation of set union. The result, as the diagram in Fig. 3 shows for X and *Z*, is a two-directional tree that contains all the events necessary to solve any conditional probability problem. Of course, as students become familiar with this method the Venn diagrams can be dropped from the tree and replaced with probability values (see Fig. 4). Consider the following problem originally presented by Tversky and Kahneman which has become something of a classic in the literature on conditional judgement (Tversky and Kahneman, 1974; Scholtz, 1987).

> An old man witnesses a hit and run accident and reports that the car involved was a blue cab (statement *"b").* There are two taxi companies in the town, the Blue *(B)* with 15 cars and the Green (G) with 85. At the trial the man's vision is tested

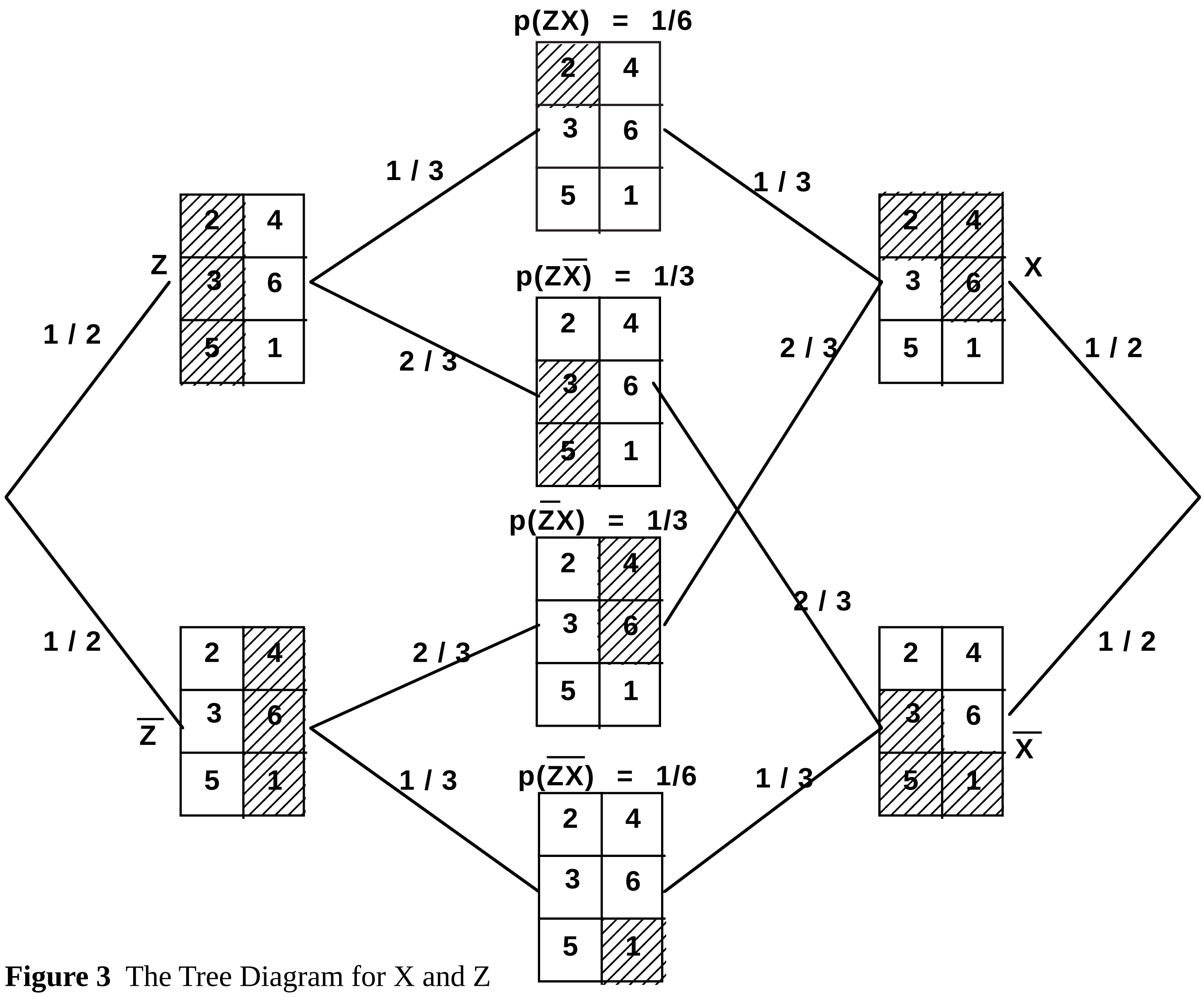
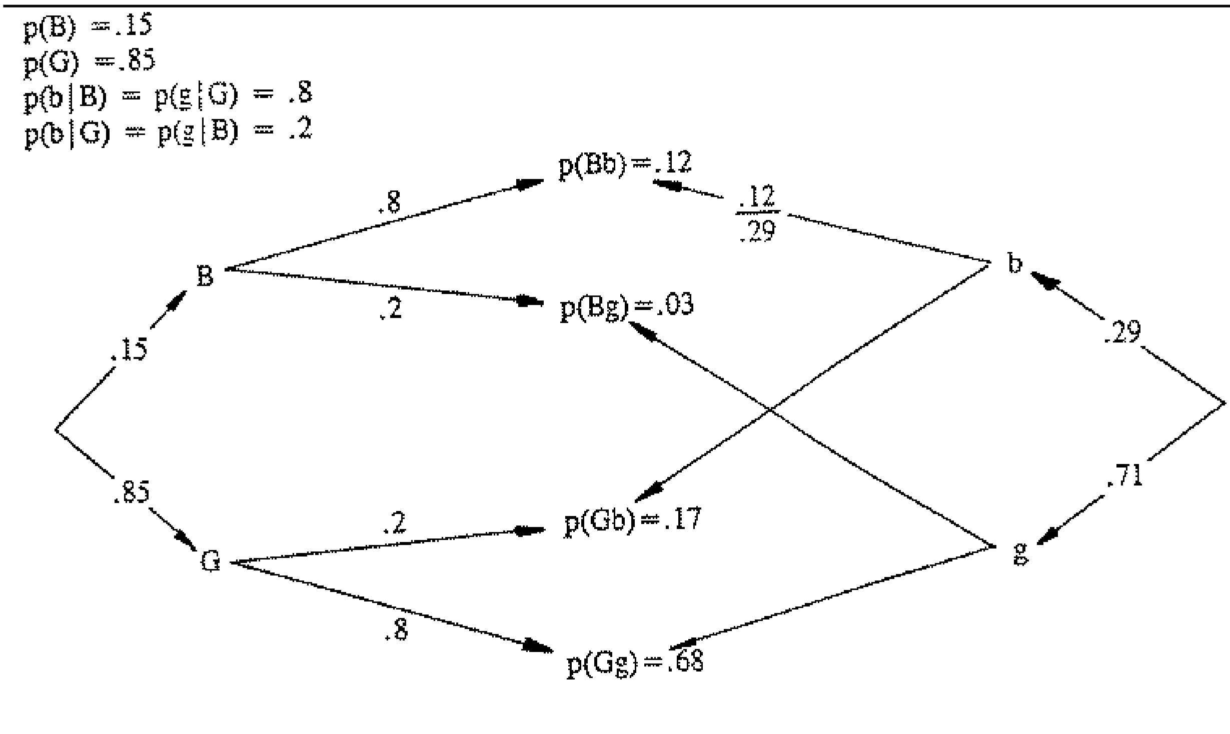


Figure 3 The Tree Diagram for X and Z



p(B|b) = .12/.29 = .41

Figure 4 The Hit and Run Affair

and he is found to be capable of identifying the correct colour cab 80% of the time (i.e. $p(b|B)$ = p(gIG) = 0.8).

Given this information Tversky and Kahneman then pose the following Bayesian question:

Q: What is the probability that a blue car was involved in the crime?

According to their report most subjects guess that the value of Q is over 1/2. In our own trials we found that undergraduate mathematics education majors were also inclined to “accept the old man’s judgement as reliable”, and assume the blue car was therefore most likely the one involved. Yet, however compelling this argument may seem, a jury which adopted the same reasoning would probably send the wrong driver to jail! For, following our argument, the probability that the blue car was involved, given the old man’s claim, is

$$p(B|b) = P(Bb) / p(b)$$

As Fig. 4 shows,

$$p(b) = p(Bb) + p(Gb) = 0.12 + 0.17 = 0.29$$

and therefore $p(B|b) = 0.12 / 0.29 = 0.41$

While Shaughnessy is surely correct to advocate simulating these frequencies in a number of trials, the simple method we propose for computing conditional probabilities appears, from our experience, to be well within the grasp of most competent students. Indeed, recognising that theory ought to follow practice, we even found that students could derive Bayes’ Theorem for themselves through the simple exercise of completing the extended tree diagram for events *A* and *B.*

◆CONCLUSION ◆

Conditional probability is a difficult topic for students to master. Often counter-intuitive, its central laws are composed of abstract terms and complex equations that do not immediately mesh with subjective intuitions of experience. If students are to acquire the mathematical skills necessary for rational judgement, teaching must focus on challenging the personal biases and cognitive heuristics identified by psychologists, and demonstrate in the most accessible way the power of probabilistic reasoning. In our discussion we suggest one graphic tool that can help in this endeavour. While the model we propose may not fit all conditional probability problems, we believe that, combined with the kind of empirical explorations discussed by Shaughnessy, it can add to the teacher’s armoury of methods for helping students come to grips with dependent and independent events, normative rules for combined events, and the logic of Bayesian questions.

References

Gardner, H. (1991). *The Unschooled Mind.* New York: Basic Books.

Grey, D. R., Holmes, P., Barnett, V. and Constable, G. M. (Eds.) (1983). *Proceedings of the First International Conference on Teaching Statistics.* Sheffield: Teaching Statistics Trust.

Falk, R. (1994). Inference Under Uncertainty Via Conditional Probabilities. *Studies of Mathematics Education: Vol 7. Teaching Statistics in Schools.* Paris: UNESCO.

Fischbein, E. (1987). *Intuition in Science and Mathematics.* Dordecht, The Netherlands: Reidel.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgement Under Uncertainty: Heuristic and Biases.* Cambridge: Cambridge University Press.

Piattelli-Palmarini, M. (1994). *Inevitable Illusions.* New York: John Wiley & Sons.

Scholtz, R. (1987). *Cognitive Strategies in Stochastic Thinking.* Dordecht, The Netherlands: Reidel

. Shaughnessy, J. M. (1992). Research in Probability and Statistics: Reflections and Directions. *Handbook of Research on Teaching Mathematics and Learning* D. A. Grouws (Ed.). New York: Macmillan.

Tversky, A. and Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science* 185, 1124-1131.