Diagrams for difficult problems in probability

PETER C-H. CHENG and NIGEL G. PITT
School of Psychology, University Park, Nottingham, NG7 2RD, U.K.
Email: peter.cheng@nottingham.ac.uk

We have developed a novel diagrammatic approach for understanding and teaching probability theory – Probability Space diagrams [1]. Our studies of learning and instruction with Probability Space (PS) diagrams have demonstrated that they can significantly enhance students' conceptual understanding. This article illustrates the utility of PS diagrams by applying them to the explanation of some difficult concepts and notoriously counter-intuitive problems in probability. We first outline the nature of the system.
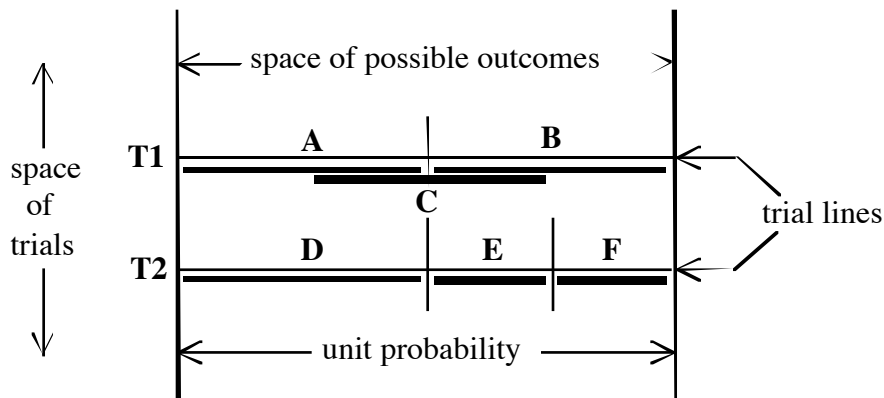

FIGURE 1. Definition of PS diagrams. Outcome segments represented by the lines A-F.

Probability Space Diagrams

Figure 1 defines the basic features of a Probability Space diagram. A PS diagram consists of a vertical strip of the plane with a constant width, the *probability space*. The width of the probability space is taken to represent either unit probability or the number of entities in the universe of interest. Each horizontal *trial line* across the space represents a trial consisting of a number of possible outcomes. Each subdivision of the trial line represents a possible outcome of that trial, an *outcome segment* (e.g. in Figure 1, *A*, *B* and *C* are possible outcomes in trial *T1*). The length of each outcome segment represents the probability of the occurrence of that outcome, or its relative frequency (e.g. P(*A*) = P(*B*) = P(*C*)). Spatial relations amongst the outcome segments capture the set theoretic relations amongst the possible outcomes; overlapping outcome segments show the intersection of the outcomes (e.g. *A* and *B* are mutually exclusive, *C* intersects both *A* and *B*).

Successive trial lines arranged vertically within the probability space represent other trials. Spatial relations amongst the outcome segments across different trial lines are used to encode dependency relations amongst outcomes in different trials (e.g. in trial *T2* outcomes *E* and *F* are possible when the outcome of trial *T1* is *B*, otherwise outcome *D* occurs given outcome *A*).

Basic interpretations of PS diagrams

Figure 2 shows all the possible outcomes when a biased die is thrown once in which the probability of a 6 is twice as likely as each of the other equi-probable outcomes. It is obvious from the diagram that $P(6) = 2/7$. The diagrammatic interpretation of the law of complementary events is simply the width of the probability space minus the length of the outcome segment of interest (e.g. $P(\text{not } 6) = 1-P(6) = 5/7$). In addition to the magnitudes for probability the diagram also shows the odds. The ratio of the length of an outcome to its complement represents the odds of the outcome (e.g. $\text{Odds}(6) = P(6)/P(\text{not } 6) = 2/5$).
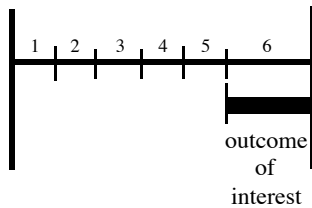
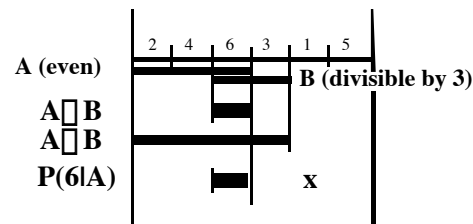FIGURE 2. PS diagram for the toss of a biased die.

FIGURE 3. Some of the relations among the outcomes of even numbers and numbers divisible by three.

Figure 3 shows a PS diagram that emphasises two outcomes (getting an even number, *A*, getting a number divisible by 3, *B*) that are different conjunctions of the possible numbers when a fair die is thrown. The diagram illustrates how the probability of the union of two outcomes is equal to the sum of their individual probabilities minus their intersection ($P(A \cup B)=P(A)+P(B)-P(A \cap B)$). Further, by considering just a sub-space of the whole probability space conditional probabilities can be considered. For example, the conditional probability of getting a 6 given an even number is represented by considering just the left side of the diagram (i.e. $P(6|\text{even} = 1/3)$). The cross indicates the rest of the probability space is being ignored.

Independent outcomes and trials

Independence is an important concept in probability theory. The following problem is one in which it is essential to assume the independence of successive similar trials.

> **Basketball problem.** A basketball player is about to shoot two free throws that may win the game. Over the season she has made 2/3 of her free throws. What is the probability that she makes both free throws?

Figure 4 shows the generic PS diagram solution to the problem, where: P(A) is the probability of a basket (2/3); P(B) is the probability of missing; a is the length of the outcome segment for A in the first trial line, and b for B, respectively. For situations in which successive trials are independent of each other, all the possible outcomes of the second trial are drawn to scale under each of the individual outcomes of the first trial. In effect, the outcomes of the first trial act as conditionalising probability sub-spaces for the second trial. This scaling of outcomes in successive trials encodes the multiplication rule (e.g. P(A and B) = P(A)P(B) = $a{\cdot}b$). Notice how the geometric scaling of the outcomes preserves the overall prior probability of each outcome in the second trial (e.g., $a = a{\cdot}a + a{\cdot}b = a(a+b) = a{\cdot}1$). Hence, the solution to the basketball problem is given by the length of the outcome segment for two successive good throws (i.e. P(A and A) = $a{\cdot}a$ = 4/9).
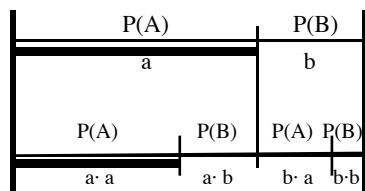
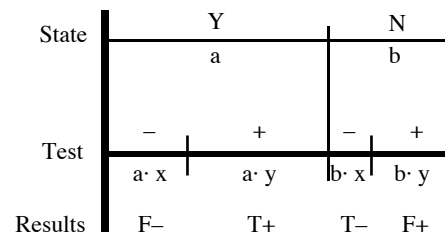**FIGURE 4.** Independent trials with similar outcomes.

FIGURE 5. Test situations exemplifying independent trials with different outcomes.

The outcomes of each trial need not be the same. A common and important example of different trial situations is where a test is conducted for the occurrence, or not, of a particular state of affairs; for example, whether a patient has a particular disease or the validity of a scientific hypothesis. If the accuracy of testing does not vary with the state then the trials are independent. Figure 5 shows the possible positive or negative outcomes of the test ('+' or '–') for both the occurrence and non-occurrence of the state of affairs (*Y* or *N*). The diagram makes explicit the need to address false positives (*F+*) and false negatives (*F*–) alongside true positives *(T+)* and true negatives (*T*-) in such test situations. By explicitly distinguishing all the alternatives, this PS diagram model of test situations may alleviate the common errors in diagnostic reasoning that arise from the neglect of false negatives and positives.

Combining independence and conditionals

When problems mix the concept of independence with the concept of conditionals, which are superficially contradictory notions, the scope for confusion is substantial. Consider this problem.

> **Cab problem** A cab was involved in a hit and run accident at night. There are two cab companies that operate in the city, a Blue Cab company and a Green Cab company. It is known that 85 cabs in the city are green and 15 are blue. A witness at the scene identified the cab involved in the accident as a blue cab. This witness was tested under similar visibility conditions, and made correct colour identification in

80% of the trial instances. What is the probability that the cab involved in the accident was a blue cab rather than a green one?

The problem requires students to not only recognise that the situation calls for computing conditional probabilities but, as Shaughnessy observes, 'there is a good deal of cognitive strain involved in reading the problem, and keeping everything straight; it is difficult for students to interpret exactly what they are being asked to do' [2 p471]. It is thus hardly surprising that problem-solvers often ignore the base rate information [3].

Figure 6 provides the PS diagram solution to the problem, which conforms to the generic model test situations (Figure 5). The diagram disentangles the consideration of conditional relations from the underlying situation involving independent trials. The conditionalising sub-space is given by the cases where the test yields a positive identification (T+ and F+; i.e., 0.12 + 0.17 = 0.29). The answer is given by the true positive as a proportion of all positives (i.e. 0.12/0.29 = 0.41).
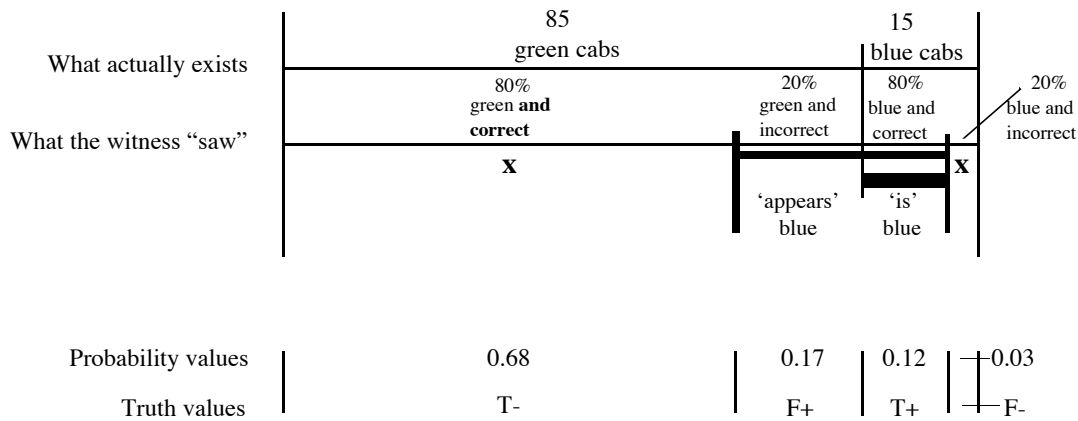

FIGURE 6. Solution to the cab problem including probability values and truth values.

Dependent outcomes

When outcomes across trials are not independent counter-intuitive problems can arise even when additional complexities such as conditionals are absent. There are two basic ways in which a trial may depend on the outcomes of the previous trial. First, the set of outcomes may be the same in the second trial, but the individual probabilities may vary according to the particular outcomes of the previous trial. For example, this would be the case in test situations if the test accuracy depended on the actual state (e.g. in the cab problem a blue cab might be more accurately identified than a green one). Second, the sets of outcomes in the second trial may vary with the particular outcomes in the first trial. For instance, this happens when different things are done on each of the outcomes of the toss of a coin, say, throwing a die given a head and cutting a pack of cards given a tail.

A fine example of the first type of dependency yielding a complex problem is Simpson's paradox [4].

Simpson's Paradox. A rookie baseball player, R, believes he will be selected for the game because his batting average is 0.224 compared to an old timer on the team, OT, whose average is 0.186. However, R is not chosen for the first game because the opposition have a right-handed pitcher and OT's batting average, against right-handed pitchers, compared to R's, is the greater, 0.179 to 0.162, respectively. In the next game the opposing side has a left-handed pitcher, so R naturally expects to be chosen, but he is not. Again, OT's batting average is 0.560 whilst R's is 0.322 against left-handed pitchers. How can this be explained?

The PS diagram for this problem is shown in Figure 7, which for the sake of clarity shows only hits with exaggerated outcome segments. Although R's overall average ($c+d$) is greater than that of OT ($a+b$), the individual percentages against right-handed and left-handed pitchers are indeed in OT's favour ($b/R_{OT} > d/R_R$ and $a/L_{OT} > c/L_R$). This is possible only because of the very different proportions of left-handed and right-handed pitchers that R and OT have faced. In such cases, contrary to our expectations, the weighted average of the averages does not preserve the order of the individual percentages, hence the paradox.
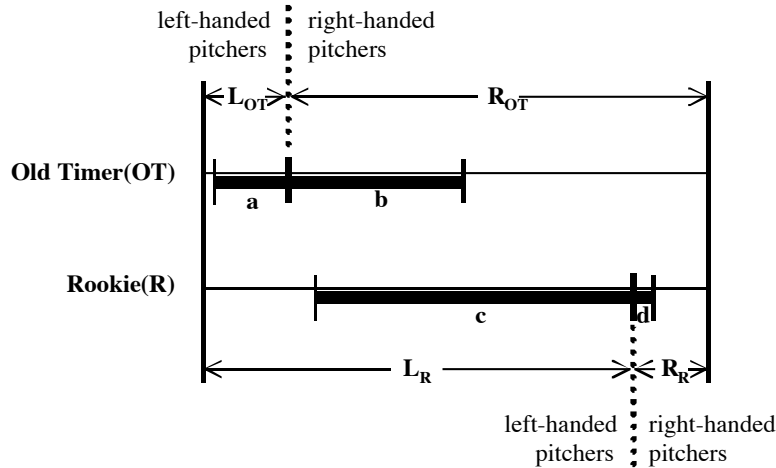

FIGURE 7. PS diagram for the explanation of Simpson's paradox

The next problem illustrates how the second form of dependency, in conjunction with a conditional relation, can result in a particularly puzzling problem.

Three Prisoners Problem. Prisoners A, B and C were in jail. Prisoner A knew that one of them was to be set free and the others were to be executed, but A didn't know who was the one to be spared. To the jailer who did know, A said, "Since two out of the three will be executed, it is certain that either B or C will be, at least. You will give me no information about my own chances if you give me the name of one man, B or C, who is going to be executed." Accepting this argument after some thinking, the jailer said, "B will be executed." Thereupon A felt happier because now either he or C would go free, so his chance had increased from 1/3 to 1/2. This prisoner's happiness may or may not have been reasonable. What do you think?

The chance of A being freed does not change. 'It is an especially good example which illustrates the drastic discrepancy between intuitive reasoning and mathematical formal reasoning about probability. In particular, it raises intriguing

questions concerning the mathematical and cognitive relevance of factors such as prior probabilities and the context in which certain information is given' [5 p1]. The PS diagram solution, Figure 8, does not only show that the answer is 1/3 but also provides an explanation of why it is so confusing.

The first line shows the prior probabilities of each prisoner being the one to be freed. The second line shows whom the jailer can say will be executed: if A is to be released then the jailer can say either 'C' or 'B' will be executed; if either C or B is to be set free then the jailer must say that the other will be executed, because the jailer will not tell A that it is he who will be released. As the jailer says B will be executed this defines a new probability sub-space for consideration; parts of the probability space set aside are marked X. Now, the probability that A is freed is given by that part of the outcome segment for A in the first line that falls within the new probability sub-space, as shown by the third line.
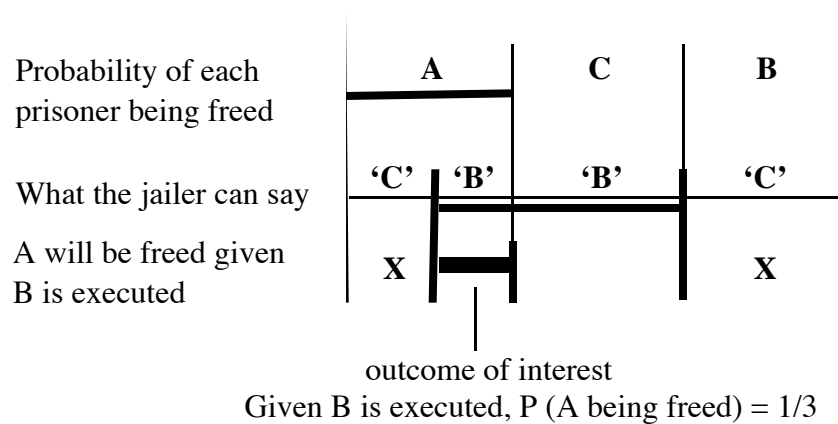

Given B is executed, P (A being freed) = 1/3

FIGURE 8. PS diagram solution to the Three Prisoners problem.

The diagram shows the two reasons why this problem is difficult. First, there are complex dependencies between who will really be freed and whom the jailer can say will be executed. Second, the conditional selects a sub-set of these dependencies that interacts with the prior probability of A being freed. The PS diagram makes that underlying logic of the problem more transparent, which is typically lost in verbal considerations of the complex relations amongst outcomes across trials.

The Monty Hall Dilemma [2] is another notoriously difficult problem that arises from complex dependencies amongst outcomes across different trials. A PS diagram solution to that problem, which also clarifies its underlying logic, can be found in Cheng [1].

Multiple trials

In problems with more than two trials the effects of independent and dependent outcomes are often cast in terms of sampling from a population with or without replacement, respectively. Further, permutations and combinations become of interest when there are multiple trials. Understanding the nature of the interaction of

sampling strategy and the grouping of outcomes provides a conceptual challenge, as illustrated in the next problem.

> **Video selection problem.** In a sale you buy three new videos. One of these, though you do not know it yet, is of poor technical quality. You choose beforehand at random the order in which you intend to play the three videos. Find the probability that the defective video is: (1) the first that you attempt to play; (2) the second that you attempt to play; (3) the third that you attempt to play.

One might think that the probability increases with successive videos in this no replacement problem, but the probability is actually the same, 1/3, in all three cases. Figure 9 shows how locally the set outcomes in successive trials differs according to the specific outcomes in the previous trial, but also that at a global level these individual outcome selections occur systematically across the diagram in such a way as to compensate for the local variations. Thus, the chances of the defective video being selected on any particular trial is constant.
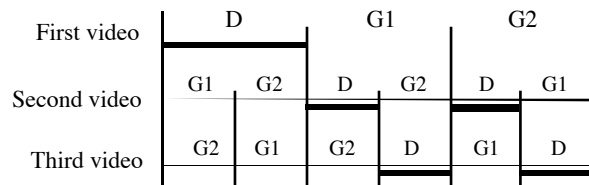

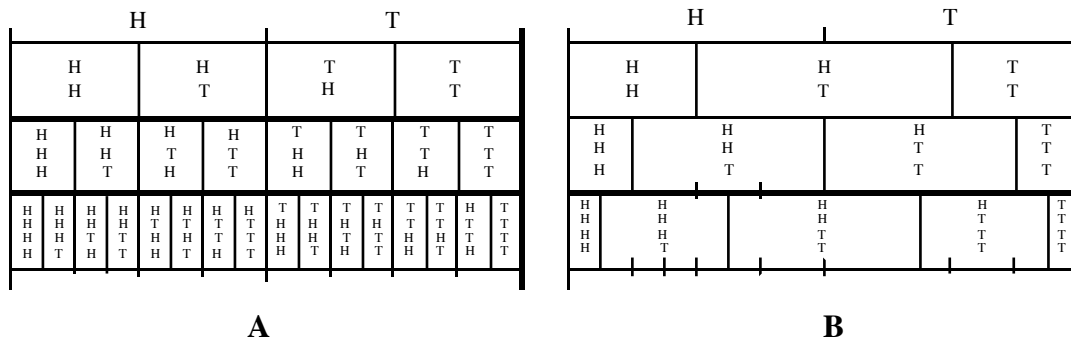**FIGURE 9.** Video selection problem. D, G1 and G2 are the defective and two good videos respectively.

A   B

FIGURE 10. Permutations (A) and combinations (B) on four tosses of a coin.

Understanding probability distributions

The nature of statistical distributions follows naturally from the consideration of multiple trials, and the Binomial distribution is often used as a bridge between the two topics. PS diagrams can be used to support this transition by providing coherent explanations of the conceptual links between probability theory and binomial distribution.

First, how does the underlying stochastic process generate the possible permutations and combinations of outcomes that make up the distribution? On successive tosses of

a coin, why does the number of permutations rise geometrically whereas the numbers of combinations increase arithmetically by one with each successive trial? Figure 10 shows PS diagrams for the two cases. Every outcome in any trial in the permutation case (e.g., HHT) contributes uniquely to two distinct new sequences (e.g., HHTH and HHTT), hence the successive doubling. In the combinations case, each new outcome (e.g., [H H T T]) receives contributions from two outcomes in the previous trial (e.g., [H H T] and [H T T]), except the pure H and T combinations. This sharing rather than division of outcomes is responsible for the unitary increase in numbers of combinations.

This sharing process also provides an explanation of the coefficients encoded in Pascal's triangle, the numbers of different ways each combination can be uniquely generated (permutations in each combination). The coefficient for a particular combination in a given trial can be found directly from Figure 10(B) by counting the number of the unit length permutation line segments.

The probability of each combination or permutation is also represented as the length of its line segment in the usual manner of PS diagrams. This remains the case when the underlying probabilities of the outcomes on each trial are unequal. The varying lengths of the lines for the different items are responsible for the skewed nature of distribution that is obtained, as illustrated in Figure 11A.
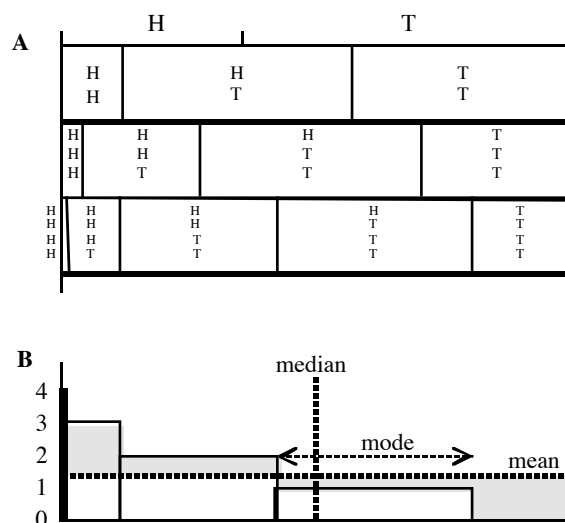

FIGURE 11. (A) PS diagram for the Binomial distribution constituted by four trials with unequal outcome probabilities. (B) Mean, median and mode numbers of Hs at the fourth trial.

A connection between the nature of the distribution and the measures of distributions can be made by extending PS diagrams. Figure 11B illustrates this for the case of four tosses of a tail biased coin, where the vertical axis is used to encode the numbers of heads. As the horizontal axis represents probability and the outcomes segments are arranged in order of increasing numbers of heads, there are simple geometric interpretations of basic measures of the distributions. The mode is the value of the bar with the greatest width. The median is the value of the bar that is cut by the

bisector of the probability space. The mean is the height of a rectangle across the whole probability space that has an area equal to the sum of the areas of all of the individual bars. The area of the bars above this mean rectangle must equal the area within the rectangle not filled by any bar, which are shown by the shaded areas in the diagram. As the distribution in Figure 11(A) is highly skewed Figure 11(B) shows that, although the median is equal to the mode, the mean is not.

Scope and limitations

The potential of PS diagrams for supporting students' conceptual understanding has been illustrated by the relatively parsimonious solutions they provide to some notoriously challenging problems. However, PS diagrams are not universally applicable to all classes of problems in such a straightforward manner. The utility of PS diagrams is diminished when there are many trials and large numbers of outcomes per trial, because the geometric increase in numbers of outcomes and the decrease in the segment sizes make drawing difficult. Thus, obtaining numbers of permutations and combinations in situations much more complex than Figure 10 is impractical. This does, nevertheless, provide a concrete illustration of the dramatic nature of combinatoric situations.

PS diagrams are effective when the situation can be modelled by the incremental construction of the diagram, which is normally the case if the probabilities for all the outcomes in at least one trial are given. They are less effective when the diagram has to be refined iteratively, which may occur when the probabilities are given as relations across outcomes in different trials. The difficulties are exacerbated in problems that require some outcome to be an extremum.

Conclusions

PS diagrams constitute a novel codification of probability theory. We have deliberately avoided the use of *event* as a term when presenting PS diagrams, because its use in conventional treatments of the topic is often ambiguous. It is regularly used to refer simultaneously to trials and to outcomes. PS diagrams were specifically designed to maintain a clear conceptual separation between the two notions by the use of orthogonal spatial dimensions to represent outcomes and trials.

PS diagrams provide a scheme in the form of the probability space. This helps to ensure that all the possible combinations of outcomes are considered, as in the case of test situations (Figure 5). PS diagrams simultaneously display the details of particular cases and show how the laws of probability, encoded as simple geometric rules, determine the specific form of the cases. The explanation of Simpson's paradox is a good example of this (Figure 7). The diagrams allow the modelling of the general problem situation to be considered independently of the interpretation of particular relations. This simplifies problems such as the cab problem (Figure 6) and the video problem (Figure 9) by decomposing their solution into distinct stages. The modelling of complex contingencies is also supported by PS diagrams which allow links between outcomes across multiple trials to be enumerated, as with the three prisoners problem (Figure 8) and in the explanation of the basis of the Binomial

distribution (Figure 10B). By reading the same diagram in different ways it is possible to contrast alternative measures of probability or different outcomes (Figure 6) or to compare relations amongst outcomes (Figure 3).

These features of PS diagrams that support the explanations of the complex problems considered here appear also to be implicated in the findings of our empirical studies of learning and instruction using PS diagrams with simpler problems [1]. In the studies the students using just the PS diagrams gained a significantly better conceptual understanding of probability theory than those using a traditional approach, which combined algebraic expressions with Venn diagrams, tree diagrams and sample space diagrams. PS diagrams apparently provide a basis for the fundamental comprehension and learning of probability theory.

Diagrammatic representations have long been used to support learning about probability. Venn diagrams and tree diagrams are well-established and geometric probability [6] and pipe diagrams [7] are more recent innovations. In contrast to these other approaches, one major benefit of Probability Space diagrams is that they constitute a single representational system that encodes the majority of the important concepts necessary for understanding probability. The examples presented illustrate how the scope of PS diagrams spans the consideration of set relations, situations with single and multiple trials, combinations and permutations, alternative probability relations such as conditional outcomes and independent trials, and various measures of distributions. The conventional approach to probability instruction requires different representations to be learnt to achieve the same coverage, which place a substantial burden on the student. The integration of all the different functions of those separate representations into a single system is why PS diagrams can provide such coherent solutions to some of the most notorious problems in probability.

References

1.   Cheng, P. C.-H. (1999). *Representational analysis and design: What makes an effective representation for learning probability theory?* (Technical Report No. 65). ESRC Centre for Research in Development, Instruction & Training

2.   Shaughnessy, J. M. (1992). Research in Probability and Statistics: Reflections and Directions. In D. A. Grouws (Eds.), *Handbook of Research on Mathematics Teaching and Learning* (pages 465-494). New York, USA: Macmillan

3.   Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.

4.   Reinhardt, H. E. (1981). Some Statistical Paradoxes. In A. P. Shulte & J. R. Smart (Eds.), *Teaching Statistics and Probability* (pages 100-109). Reston, Virginia, USA: National Council of Teachers of Mathematics.

5.   Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the "problem of the three prisoners". *Cognition,* Vol. 32, pages 1-24.

6.   Dahlke, R & Fakler, R. (1981). Geometrical probability. In A. P. Shulte & J. R. Smart (Eds.), *Teaching Statistics and Probability* (pages 143-153). Reston, Virginia, USA: National Council of Teachers of Mathematics

7.   Konold, C. (1996). Representing Probabilities with Pipe Diagrams. *The Mathematics Teacher*, Vol. 89 (May No 5) pages 378-382.