Axioms of Probability

The axioms and other basic formulas for the algebraic treatment of probability are considered.

Axiom 1: For any event, A, that is a member of the universal set, S, the probability of A, P(A), must fall in the range,

$$0 \le P(A) \le 1 \ . \qquad 3$$

Axiom 2: The probability of S is,

$$P(S)=1 \ . \qquad 4$$

Axiom 3: If A and B are disjoint events, $A \cap B$ is empty, then,

$$P(A \cup B) = P(A) + P(B) \ . \qquad 5$$

(The fourth axiom for infinite disjoint events is omitted here.)

Various propositions follow from the axioms. The probability relation between complementary events is,

$$P(\sim A) = 1 - P(A) \ . \qquad 6$$

Before discussing the other relations, derived from the axioms, the classes of probabilistic situations need to be considered.

Classes of Probabilistic Situations

Figure 3 shows a classification of the dimensions of probabilistic situations. This particular conceptualization of the general underlying structure of this topic was developed as part of the work described here. Why an existing conceptualization could not merely be adopted is discussed below and demonstrates how representations can influence the overarching conceptualization of domains. At the top level the classification distinguishes between orthogonal dimensions of <u>dependent</u> events and <u>independent</u> events. This highest level division will be considered first and followed by the lower level subdivisions.

Independent versus dependent events

The dependent dimension considers events within one trial and the independent dimension deals with events across more than one trial. On the first dimension, events are dependent because the occurrence of any event will mean logically that other events may or may not occur (e.g., an odd number thrown on one die means an even number is impossible but a prime number may be possible). On the second dimension, events in two or more trials are considered. The occurrence of an particular event, E1, in one trial will determine the set of events, S2, for consideration in the next trial (e.g., toss a coin, if a head is obtained toss the coin again, otherwise throw a die). Events in the two trials are independent in the sense that the occurrence E1 does not logically determine any relations among the possible outcomes of the events in S2, although it does determine that S2 is selected. In other words, the occurrence of E1 does not by itself provide any information about the structure of S2 (e.g., getting a head on the toss of a coin that selects the throw of a die does not influence which of the six numbers will be obtained on casting the die.)

Quite different processes are involved in independent situations compared to dependent situations. Whereas the operations for the dependent class of relations are grounded in set theory, those for independent relations are not. To make this concrete, consider the use of a Venn diagram to represent the dependent occurrence or not of an even number on the throw of a die, Even or Odd, Figure 4. But how should the diagram be modified to incorporate a second

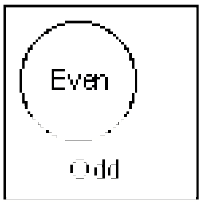

Figure 4. Venn diagram for the outcomes of throwing a die

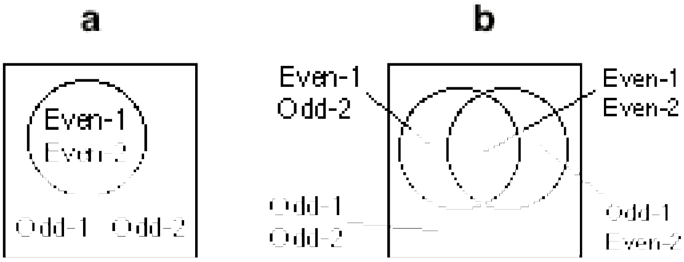

Figure 5. Failed attempts to modify Venn diagram for independent trials. (Trial numbers are indicated by suffixes on the outcome labels.)

independent throw of the die? Letting the Even circle simultaneously represent both throws will not do, Figure 5a, as the combination of an even preceding (or following) an odd number is not represented (e.g., Even-1 & Odd-2 is not represented). Drawing another circle for even numbers on the second trial that partially overlaps the first circle allows all four possible permutations of trial outcomes to be represented, Figure 5b. This, however, is also unsatisfactory, because each circle no longer simply encompasses only one type of event (e.g., just even numbers). The difference in the processes underlying independent and dependent trials is reflected in the structure of the formulas that encode the probability relations for the two classes of situations.

An examination of probability texts shows that this logical difference is not usually made explicit. The use of the symbols from set theory for 'and' and 'or' in equations for probability relations in both classes is symptomatic of this. It is hypothesized that this conflation of different types of situation is a problem for the learners of probability. PS diagrams clearly distinguish the classes, and how this problem was highlighted during the development of PS diagrams is described below.

Skemp (1986) recommends the use of different symbols to distinguish mathematical procedures that on an abstract conceptual level are equivalent but which require different types or combinations of operations to perform (e.g., addition of integers versus addition of fractions). Here, different symbols will be used for the relations of “and”, “or” and “given” in independent versus dependent situations. For dependent situations the symbols will be borrowed from set theory. For independent situations ‘&’, ‘v’ and ‘«’ will be adopted for “and”, “or” and “given”.

Dependent Situations

Dependent situations can be subdivided into those cases where the events of interest are joint or disjoint events, Figure 3. For example, in the throw of a die obtaining an odd number is disjoint from getting an even number but joint with getting a prime number.

In general, equation 5 gives the relation between the individual probabilities and the probability of the union of disjoint events. For events that are not disjoint (intersect or not mutually exclusive) the relation to the individual probabilities and the probability of their intersection is,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \; . \qquad 7$$

Equation 7 reduces to Equation 5 when the sets are disjoint.

The multiplication theorem relates conditional probability of dependent event A given event B to the joint probability of A and B,

$$P(A \cap B) = P(A|B)\, P(B) \; . \qquad 8a$$

Similarly, for B given A, the formula is,

$$P(A \cap B) = P(B|A)\, P(A) \; . \qquad 8b$$

Independent Situations

Two events, A and B, in different trials are independent when,

$$P(A\&B) = P(A)\, P(B). \qquad 9$$

The probability of A or B is given by the formula,

$$P(AvB) = P(A) + P(B) - P(A\&B) = P(A) + P(B) - P(A)\, P(B) \; . \qquad 10$$

The conditional probabilities of independent event A given event B in an earlier trial, is simply the prior probability of A, because they are independent,

$$P(A\text{«}B) = P(A) \; . \qquad 11a$$

Similarly, for B given A, the formula is,

$$P(B\text{«}A) = P(B) \; . \qquad 11b$$

The probability of A or B can be found even when A&B is not given using,

$$P(AvB) = 1 - P(\sim A\&\sim B) \; . \qquad 12$$

The dimension of independent events has sub-divisions of unlinked or linked events, Figure 3. The terms <u>linked</u> and <u>unlinked</u> event have been coined here, because probability texts do not usually discuss this sub-dimension directly, but often cover the concepts implicitly using worked examples.

Unlinked situations come in two forms. <u>Similar-unlinked</u> situations repeat the same process on successive trials, for example, the multiple toss of a coin, throwing several die, and selecting beads from a bag with replacement between selections. <u>Dissimilar-unlinked</u> situations have different processes on successive trials; for example, in the Cab problem one trial deals with the base rate of cabs and the second trial concerns the accuracy of the witness's identification of the colour of the cabs.

One reason why the Cab problem is difficult is knowing what to do with the information about the base rates, which are commonly ignored by naive and not so naive reasoners. To keep track of the possibilities of correct or incorrect identification of blue or green cabs, the use of tree diagrams, Figure 6, or contingency tables, Figure 7, are typically recommended (e.g., McColl, 1995; Shaughnessy, 1992). The cells in the table match the leaves of the tree. Although they help clarify what pair-wise calculations to do, they do not show which intermediate values to pick and what calculation to do with them to obtain the overall result.

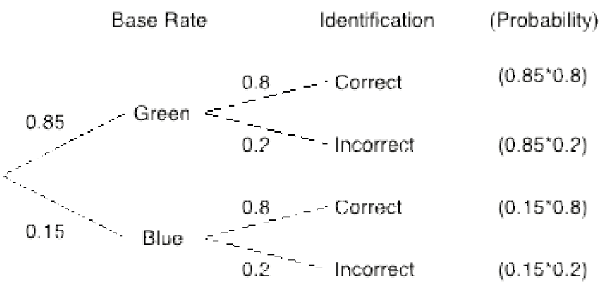

Figure 6. Tree diagram.

Witness identification (0.8 accurate)

Taxis	Blue	Green
Blue 0.15	0.12	0.03
Green 0.85	0.17	0.68
	0.29	0.71

Figure 7. Contingency table.

In linked situations the particular composition of the groups of events in a trial differ from each other and from the set of events in the preceding trial. However, which group is chosen in the second trial is determined by, or linked to, the particular event occurring on the first trial. Picking beans from a bag without replacement is a linked situation, because not replacing the beans changes the set of beans for the next trial. Monty's Dilemma includes linked independent situations. Another example is the case above, where choosing whether to toss a coin again or to throw a die is linked to the outcome of the initial toss of a coin. This example is included in Figure 3 along with examples for the other types of situation. (The 'encoding' entries are discussed below.)

Bayes' theorem provides a general approach to problems with independent situations, but it is normally introduced to advanced students and is often difficult to apply (see below).

Permutations and Combinations

Outcome tables can be used to enumerate the possible outcomes of independent situations. Figure 8 shows all 36 outcomes of throwing two dice. Each cell gives one possible outcome permutation. The other number in each cell is the sum of the numbers from each die. Such outcome tables are often used for instruction, because some events are easily read-off the diagram, such as the number of sevens along the diagonal. Similarly the numbers of combinations can be found by cutting the grid along the other diagonal and counting the cells above the line. From the structure produced the number of combinations can be seen to be the sum of the arithmetic series of integers from 1 to 6.

Table 1 gives the general formulas for the numbers of permutations or combinations of k selected objects from an initial set of n objects, for selection with and without replacement. Some effort is required to use the formulas to examine the numbers of permutations and combinations under different situations in order to gain a good qualitative appreciation of the relations among them.

Die 1 \ Die 2	1	2	3	4	5	6
1	1,1 2	1,2 3	1,3 4	1,4 5	1,5 6	1,6 7
2	2,1 3	2,2 4	2,3 5	2,4 6	2,5 7	2,6 8
3	3,1 4	3,2 5	3,3 6	3,4 7	3,5 8	3,6 9
4	4,1 5	4,2 6	4,3 7	4,4 8	4,5 9	4,6 10
5	5,1 6	5,2 7	5,3 8	5,4 9	5,5 10	5,6 11
6	6,1 7	6,2 8	6,3 9	6,4 10	6,5 11	6,6 12

Figure 8. Permutation and sum outcomes of throwing two dice.

Table 1.
Formulas for Numbers of Permutations and Combinations.

	Independent unlinked— repetition or replacement between selections	Independent linked— no repetitions or no replacement between selections
Permutations	n^k	$\frac{k!}{(n - k)!}$
Combinations	$\frac{(n + k - 1)!}{k!(n - 1)!}$	$\frac{n!}{k!(n - k)!}$

Odds

Probabilities may be expressed in the form of odds, which are ratios of chances of an event happening to it not happening. The relation of the probability to odds for and against event A is given by the two formulas, respectively;

$$P(A) = \frac{\text{odds}(A)}{1 + \text{odds}(A)} , \qquad 13$$

and

$$P(\sim A) = \frac{1}{1 + \text{odds}(A)} . \qquad 14$$

Odds sometimes provide a convenient way to express the probabilistic relations between pairs of events, such as the chances of a hypothesis being correct given certain evidence. The likelihood ratios of necessity and sufficiency are useful ways to express the possible relations among hypotheses and evidence (e.g., Giarrantano and Riley, 1989).

Knowledge Requirements of the Domain

What does someone need to know to have a good understanding of probability? All of the following things have a role, to a greater or lesser extent.

Mastery of the principles and laws of probability is certainly essential, which in terms of an algebraic interpretation of the domain means having access to at least the equations above and knowing when and how to apply them to different situations. Knowing prototypical and extreme examples will aid comprehension of a domain by providing set points of reference against which comparisons can be made in the interpretation of novel cases. Arguably, a coherent network of such cases is the foundation of conceptual expertise. Further, such cases may serve as cues for the selection of appropriate formulas to use in modelling situations. For example, throwing two dice is a common example, which can be used as an analogue to some other situation to help identify that situation as an independent unlinked process, and so aid ones recall of formulas that are applicable (Equations 9 to 12).

There are many different levels and perspectives in the domain that need to be interrelated. Underpinning much of probability is set theory. Chance events may be considered in terms of probabilities or odds and related to relative frequencies in cases with finite samples. For repeated processes, samples may be considered in terms of permutations or combinations and related to the probability of particular sequences or to the overall numbers of sets of events.

Ideally, it is desirable for a learner to obtain a correct, complete and coherent network of concepts for the domain. Completeness is near impossible given all that is known about probability and the limited time that can be devoted to any one topic in mathematics classes. A more realistic alternative is to provide the learner with sufficient knowledge so that they can fill in the conceptual gaps for themselves as required. To be able to extend their own knowledge learners will need some understanding of the different methods and approaches that can be used for reasoning about probabilistic situations. Knowing when and how to use certain techniques is important for successful modelling. For example, given situations with large samples, it is better to use the formulas in Table 1 to compute the total numbers of permutations or combinations, rather than to attempt to enumerate and count every subset.

Learning how to deal with the complex interactions that can occur in probability problems is an important part of learning the domain. The Cab problem is more complex than naive reasoners initially imagine and Monty's dilemma is complex irrespective of one's experience in probability. In addition to competence in algebra, skill at using the other "tools of the trade" is particularly necessary for complex situations. The ability to generate outcome tables as required (e.g., Figure 8), and to use tree diagrams and contingency tables when needed (e.g., Figures 6 and 7), are components of competence in this domain.

For a really robust knowledge of the domain, an understanding of the pitfalls of the topic, the common misconceptions and possible misinterpretations, may be useful. Such knowledge can be used to check for errors and more importantly can help to delimit the boundaries of specific concepts or procedures. This may assist the learner developing a sense of the global structure or topology of the domain and thus provide a better foundation for a correct, complete and coherent conceptual network. No general claim is being made about the importance of the role of errors in learning, but it is argued that knowing about the areas of common conceptual difficulties can shed some light on the concepts themselves. This is, in particular, with regard to the limitations of the representational systems used for a domain. For example, algebraic slips are easy to make when working with equations. Learners can be made aware of some of the common hazards and can be taught to spot when errors have occurred by checking the dimensional consistency of the units of each the terms in an equation (e.g., in an equation for energy conservation all the terms should be in Joules [kg m^2/s^2]). This will provide learners with a sense that there is an additional level of constraints that operates within the laws of scientific and mathematics domains (Krauss, 1994). In probability theory an awareness that probability terms are dimensionless, whilst terms in relative frequency expressions are numbers of things, places some overarching constraint on what are legal operations on expressions for each type of measure of chance.

The traditional algebraic approach to learning probability can support most of these requirements, but it does not do so particularly well. The apparent limitations are illustrated in the comparisons with PS diagrams, below.

DEVELOPMENT GUIDELINES FOR EFFECTIVE REPRESENTATIONS

This section introduces the six characteristics of effective representations inferred from the studies on LEDs, as summarized above. These characteristics are general "systemic" features that should be considered in the broad context of the knowledge requirements of the domain and the overall structure of the representation. They are being adopted as guidelines for the development of representations, and their use in the design of the new probability representation is an indication that they are sufficiently well specified to be of utility. The characteristics are grouped according to whether they (1) deal with the encoding of information and relations in the expressions of representations (semantic transparency) or (2) concern the use and usability of representations (plastic generativity).

Semantic Transparency

This set of characteristics/guidelines are things that may make representations better able to support learners' development of a coherent, comprehensible and memorable network of concepts. They identify the features of a representation that aim to make the nature of the domain apparent in the structure of the representation itself, rather than obscuring the domain behind notational formats that are largely arbitrary with respect to the nature of the domain. These characteristics/guidelines consider relations between the "static" structure of expressions, or sets of expressions, in the representation and the major conceptual distinctions, invariants or underlying regularities of the domain.

G1 Integration of levels of abstraction

Domains have information at different levels of abstraction. At the most abstract there are the general laws or relations that cover the whole domain and to a large extent help to define the scope of a domain. A case in point are the axioms and laws of probability. At the concrete level there are descriptions of particular phenomena or sets of data pertaining to specific cases; such as the situation in the Cab problem. Levels of abstraction are integrated in a representation when the information at these different levels is tightly coordinated within the expressions of the representations. A representation that has expressions to encode the laws of a domain and separate expressions for sets of data pertaining to specific cases does not have well integrated levels of abstraction.

It is hypothesized that a good representation for conceptual learning should have well integrated levels of abstraction. It should support a close conceptual connection between the cases and the laws of a domain. Such a representation will likely have simple rules governing the structure of its expressions that encode the laws whilst simultaneously presenting data for possible cases in the domain. Better conceptual understanding may result, because the representation may reveal how the laws of the domain constrain the configuration of individual cases and how the laws determine the overall topology of the space of possible configurations. With a poorly integrated representation the onus is on the learner to deliberately make these conceptual connections for themselves.

For example, a good explanation of the Cab problem should show how the interaction between the base rate information and the measures of the witness' accuracy determine the underlying structure of the problem under the appropriate probability relations, and how this yields the actual outcome probability value.

White (1993) also considers that the conceptual gulf between abstract laws and concrete cases is a general problem in the learning of science. However, she considers that the problem should be addressed by the introduction of representations at an intermediate level of abstraction to bridge conceptual gulf. This contrasts with the claim here that it may be better to devise a single representation that reduces the conceptual distance between the levels of abstraction, rather than attempt to span the gulf between the levels by adding more representational machinery.

G2 Combining globally homogenous and locally heterogeneous representation of concepts

An effective representation should, on the one hand, support a unified overall conceptualization of the domain — a globally homogenous representation of concepts. On the other hand, the representation should not obscure conceptual differences where they naturally exist, over and above any universal invariants of the domain — a locally heterogeneous representation of concepts.

To be globally homogenous a representation should have a general scheme applicable to all its expressions that encodes the deep regularities of the domain. In some form, the general notational structure of its expressions should reflect the inherent structure of the domain. To be locally heterogeneous a representation should have distinctive notational features, at a detailed level, that are consistent with differences which exist between particular concepts. Although it may seem contradictory for a representation to be both homogenous and heterogeneous, this may be achieved by encoding global and local requirements using different structural features of the same expressions.

It is claimed that globally homogenous but locally heterogeneous representations may enhance learning in various ways. The overarching representational scheme aims to provide a rational framework that learners may use to recall and examine individual concepts and to constrain the exploration of the relations among those concepts. The similarity of expressions may show learners that the concepts are related according to certain underlying invariants that hold over the whole domain. The individual features of expressions supports learning by attempting to make a clear distinction between concepts that are, at one level, quite distinct but nevertheless closely related. The likelihood of over-generalizing a concept or unnecessarily restricting the scope of a relation may both be reduced in representations that are globally homogenous but locally heterogeneous. Further, if the expressions of a representation are relatively simple, but visually distinct, they may stand as memorable icons for the concepts they represent.

For example, it is desirable for a representation for probability theory to be globally homogenous with respect to the axioms of probability and the general propositions that immediately follow from them, Equations 3 to 6, because they are universally applicable in all probabilistic situations. It is desirable for the same representation to be locally heterogeneous with respect to the various dimensions of probabilistic situations identified in Figure 3.

The notion of locally heterogeneous representations of concepts is related to Zhang and Norman's (1994a) ideas about the external separability of dimensions of information in representations. This is considered in detail in the discussion.

G3 Integration of perspectives

Like most scientific and mathematical domains, there are different perspectives that can be adopted when dealing with probability. At least three levels exist for probability.

<u>Ontological level</u>: The first level includes the perspectives of set theory and probability theory, which are distinct ontologies ranging over different elementary entities and relations. Set theory covers discrete objects and membership of groups, whilst probability theory also deals with continuous measures of the chance or likelihood of possible events. The probability ontology can be further divided into aleotoric/frequentist (statistical or stable relative frequency) versus epistemic/Bayesian (degree of belief) ontologies, as argued for by philosophers at length (Hacking, 1975).

<u>Alternative measures</u>. The second level concerns the different ways measures of probability can be expressed, sometimes independent of the assumed underlying ontology. These include: quantities of objects, odds, and probability <u>per se</u>. For example, in the Cab problem, we are told the actual proportion of cabs in the city (quantities) and the accuracy of the witness as a percentage (probability).

<u>Viewpoints</u>: The third level deals with different points of view that may be taken within particular situations. For example, in some problems similar to Monty's dilemma it is sometimes easier to determine the probability of outcomes by considering the final outcomes and working backwards through the contingencies to the initial conditions, rather than by following the temporal sequence of trials as presented in the problem statement. Different viewpoints of the relations that hold in a particular situation may also be considered. For instance, one may view the probability of a given event, P(A), in terms of the probability under the viewpoint of the non occurrence of that event, 1-P(~A). Similarly, each side of the formulas

for de Morgan's laws, Equations 2a and 2b, can be taken as alternative views of the same overall state.

It is hypothesized that the better a representation (or set of representations) supports the integration or close co-ordination of the perspectives within and between the three levels, the easier the system may be for reasoning about, and learning in, the domain. Close integration of expressions may aid the comprehension of concepts by making available relations from other perspectives that will place mutual constraints on the possible interpretation of the concepts.

Plastic Generativity

Manipulating representations to make inferences is intrinsic to learning in scientific and mathematical domains. It is rare for solutions to problems to be found without some algorithmic work to generate new expressions that meet the requirements of the given problem situations. The Cab problem and Monty’s dilemma cannot be solved by merely substituting the given values into the formulas for the laws of probability theory. A good representation will support the user in reasoning things out for themselves by modelling different cases or concepts through the derivation of appropriate expressions. Good representations for learning must, therefore, be generative — not mere displays that present sets of data or fixed visualizations of particular relations.

The set of characteristics/guidelines in this section consider the use and usability of the representation for making inferences, solving problems and exploring the conceptual structure of the domain. They concern the "dynamic" transformation of expressions. In this section, and beyond, the term <u>operation</u> is used to denote an individual action that in one step modifies an expression, usually according to a syntactic rule (e.g., canceling a term from both sides of an algebraic equation, or adding a circle to a Venn diagram). The term <u>procedure</u> is taken to mean sequences of operations to modify expressions or to generate new expressions that meet particular sub-goals generated during problem solving.

For the purpose of exposition, an analogy to the role of representations in reasoning about a domain is made to the role of materials in the construction of physical models. Both are mediums of expression, either of physical form or conceptual structure. Plastics (polymers) are particularly suitable materials for modelling, because they can be moulded and possess certain desirable properties that facilitate the modelling process. In the same vein representations must be generative and possess characteristics that support conceptual modelling that are counterparts to the desirable physical properties of plastics. The characteristics to be considered in turn are: G4, malleability; G5, compact procedures; G6, uniform procedures.

G4 Malleable representations

An important property of many plastics that makes them ideal for modelling is their malleability. They have a degree of flexibility combined with a degree of rigidity, which makes them capable of being moulded yet able retain the desired form without external support. On the one hand, a material that is too rigid will fracture rather than be moulded (e.g., chalk). But, on the other hand, the material should not be so fluid that it flows in arbitrary directions whilst being worked upon, as this too will make it impossible to form (e.g., syrup).

It is hypothesized that a good representation for conceptual learning should be malleable, that is have expressions that are neither too rigid nor to fluid. Under a general information processing conceptualization, this characteristic may be considered in terms of the mapping between the state space of expressions defined by the representation and the concepts in the domain. A rigid representation is one in which it is impossible to navigate directly from a node (an expression) for one concept to a node for another concept, because there are no procedures in the representation itself that will allow expressions to be modified to attain the goal expression. In such cases, a switch to another representation is required. A highly fluid representation has a state space that is highly branching, so that many arbitrary routes through the space can be followed, but few of them lead to the node for the desired concept.

A rigid representation may be of limited utility in the broad context of learning, because it will be hard to generate new expressions. Charts and graphs are rigid representations, as they present information about a particular case but they cannot, in and of themselves, be modified to represent other cases without additional information being supplied. A rigid representation may impose costs in the form of additional representations to be mastered, to carry out the inferential work not supported by the rigid representation.

Similarly, a fluid representation may allow many different expressions to be easily generated but most of these expressions will not be directly meaningful with respect to the cases being considered. The size of the potential search space of expressions will be large. The notation systems for set theory and probability theory are examples of fluid representations, which can be arbitrarily manipulated to produce a great many relatively meaningless formulas. Knowledge about the representation itself, in the form of strategies to direct search down profitable paths, is necessary for representations at this end of the spectrum. Mastery of such strategies may be a substantial overhead to learning about the domain.

Supplementary representations may be used as a means to help make the choice of the right path by managing the information that is needed to solve a problem; for example, the tree diagram or the contingency table used in the solution to the Cab problem. The need for supplementary representations or representation specific strategies to deal with the disadvantages of rigid and fluid representations is an additional cost to the learning of a particular domain. Thus, malleable representations that occupy the middle ground between rigid and fluid ends of the spectrum are expected to be most effective for learning.

This characteristic is related to Stenning and Oberlander's (1995) claims about the benefits of diagrammatic representations in relation to their degree of expressivity. They argue that diagrams are limited abstraction representational systems, whose expressiveness is neither too great nor too restricted. Such representations allow a finite, but not a unitary, set of models to be inferred. The notion of malleable representations differs from Stenning and Oberlander's ideas about expressivity in that the scope of this characteristic is limited to the syntactic nature of representations. Semantic considerations are separately covered by the first three characteristics above.

G5 Uniform procedures

This characteristic, and the next, consider the complexity of the procedures for the representations of a domain. It is hypothesized that the more complex the procedures the more difficult the representation will make conceptual learning in the domain. First, there is a cost to acquiring the skills and knowledge to manipulate the representation, over and above the development of an understanding of the conceptual content of the domain. Once acquired, complex procedures still pose a problem, simply because they make the exploration of the domain more demanding. The procedures may be complex in two general senses: (G5) there may be numerous and diverse procedures, or (G6) individual procedures may involve long and tortuous sequences of operations. The formal nature and implication of these two types of complexity are considered in this section and the next.

Considering our plastic analogy again, plastics for modelling usually have a uniform consistency for good reasons. A material that is not uniform (e.g., pebbles in clay) will be hard to use, because the moulding process will not only have to achieve the desired shape, but will also need to accommodate any local variations in the consistency of the material. More skill and special techniques are needed.

For a particular domain, there will be expressions for each of the concepts and procedures to modify one expression into another in order to examine the relations between concepts. The representations for a domain will be more complex in procedural terms, if more types of processes are typically needed to transform expressions for the same set of concepts. By the same token, representations will be more uniform when most of the procedures are constituted by sequences of operations with common patterns. Different representations, by definition, have different operators, so a conceptualization of a domain that requires multiple representations will necessarily be less uniform than a conceptualization that employs just one representation. With multiple representations, additional procedures that link the representations together, by mapping information between them, are also required. An individual representation may be more or less uniform depending on the number of different techniques that are needed to use the representation. Venn diagrams have a small set of techniques for their construction and modification, largely based on the drawing and labeling of intersecting shapes. In contrast, for the laws of probability a relatively large number of algebraic techniques must be employed, including: simplification of formulas by eliminating terms; isolating particular terms on the left hand side of equations; elimination of variables by substitution of formulas into each other; solution of simultaneous equations; and so forth.

A representation which has largely uniform procedures is likely to be better for learning, for at least two reasons. First, there are simply fewer procedures to learn. Procedures mastered under one part of the conceptual space will be available for the exploration of other concepts of the domain, without the need to acquire additional procedures. Second, uniform procedures will support a more coherent approach to modelling the domain, so that knowing what procedure to use on a particular problem will be more straightforward. Thus, more attention may be paid to the relations among concepts, rather than expending effort to merely find appropriate ways to derive expressions.

Cheng (1999b) describes a representation for a domain where quite different procedures are required for the same class of problems when situations differ only in detail. For some problems a simple stepwise decompositional approach is possible but for others the best approach is to set up and solve simultaneous equations. The idiosyncratic need for different procedures is an additional obstacle for learners of that domain. If different procedures are needed for the same class of situations then learners may make unwarranted conceptual disassociations despite the underlying relations being the same. This is a particular concern with representations that do not support the combination of globally homogenous and locally heterogeneous representation of concepts, because the structure of expressions for different concepts will not indicate that such a conceptual disassociation is not appropriate.

G6 Compact procedures

Carrying the plastic analogy a little further, a model will require more time and effort to produce when many different processes are used. For example, it is inefficient to sculpt a model by laboriously chipping lumps from a solid block, carving off thin layers and finally sanding to achieve a smooth finish. It is better to inject the plastic directly into a mould, a process with just one step. Mistakes are more likely in the first approach, because more actions have to be carried out and each one must be performed with some precision.

This second form of complexity of procedures is concerned with the inherent complexity of procedures themselves. A procedure will be complex if a long and tortuous path of operations is required to transform an expression into the desired form. Conversely, a compact procedure will consist of a few operations, with the generation of a small number of intermediate expressions on the way to the target form.

It is proposed that a representation that is complex to manipulate, in this sense, is likely to be a barrier to learning for at least three reasons. First, long procedures will, quite simply, be more difficult to learn than more compact procedures, because there are more steps and a greater variety of operations to be carried out in a specific order. Second, the chance of making errors in the transformations of the expressions is less for compact procedures as there are few places to make syntactic slips or misinterpret expressions. When the error rate is high the opportunity for conceptual learning is diminished, because fewer correct expressions for different concepts can be examined and the incorrect ones will cause confusion. Third, users of the representation will be less able to attend to information that is most relevant to the learning goals of a particular exercise when substantial amounts of distracting syntactic work is necessary.

That completes the description of the characteristics of effective representations.

DESIGN CRITERIA FOR LEDS

The guidelines are proposals about the general characteristics of effective representations, but they are too general to be directly used in the detailed design of a representation. From the many examples of good and poor LEDs that have been studied, the following five criteria for effective LEDs are proposed (Cheng, 1999b), which allow new representations to be designed that are likely to meet the guidelines.

- Ca) There should be simple and clear mappings from domain properties to elements of the diagram (e.g., preserving a linear relation between magnitudes of probabilities and element size; not having a diagrammatic element for an event appear more than once in the LED).
- Cb) The diagrammatic constraints encoding the laws should be simple (i.e., use simple geometric, topological and spatial relations) and should require minimal propositional constraints to help encode relations.
- Cc) Different cases should be clearly distinguished as different configurations of the diagrams (e.g., probabilities of zero and unity should be limiting cases of the LED; different diagrammatic structures for each of the classes in Figure 3).
- Cd) The criteria Ca-Cc apply to the different levels of the domain, such as relations among the variables within components and also the interactions among components (e.g., probabilities of individual events and overall probabilities of combinations of events consisting of sets of permutations).
- Ce) The diagrammatic constraints for the different levels should be compatible, such that the rules for interactions between levels are consistent with the levels and correctly encode the nature of the interaction (e.g., the interpretation of an element or configuration of the diagram standing for an event or relation should be the same at both levels).

The criteria are related so it will be necessary to make a trade-off amongst them. For instance, it may be worthwhile allowing the constraints encoding the laws (Cb) to involve a few complications if this means that the constraints across different levels in the domain (Ce/Ce) can be made compatible in a particularly straightforward manner.

PS diagrams were designed to meet these criteria as far as possible. The following section introduces the system.

PROBABILITY SPACE DIAGRAMS — PS DIAGRAMS

Probability space diagrams use geometric and spatial diagrammatic constraints to encode the laws of probability. Like Law Encoding Diagrams in general (Cheng, 1996a. 1996b), each instantiation, or drawing, of a PS diagram represents a particular case or probabilistic situation. All PS diagrams use physical space as a direct visual analogue for the "space" of possible events and their probabilities, hence the name Probability Space diagrams. PS diagrams are typically bounded by a pair of (faint) vertical parallel lines that delimits the space for that diagram.

Visual representations of probability have been proposed previously, but they have been limited to basic situations (e.g., Armstrong, 1981; Dahlke & Kakler, 1981) or restricted to particular problem types (e.g., Ichikawa, 1989; Gigerenzer & Hoffrage, 1995). PS diagrams have been designed to cover the full range of probability situations and problem types. For set theory, PS diagrams use subdivisions of the space to represent subsets. For probability theory the interpretation uses the size of the divisions to represent magnitudes of probability, odds or relative frequency. For the main classes of probability situations (Figure 3) there are different configurations of PS diagrams. Each class of diagrammatic configuration is defined by the particular operations that encode specific relations. These aspects of PS diagrams are considered in turn.

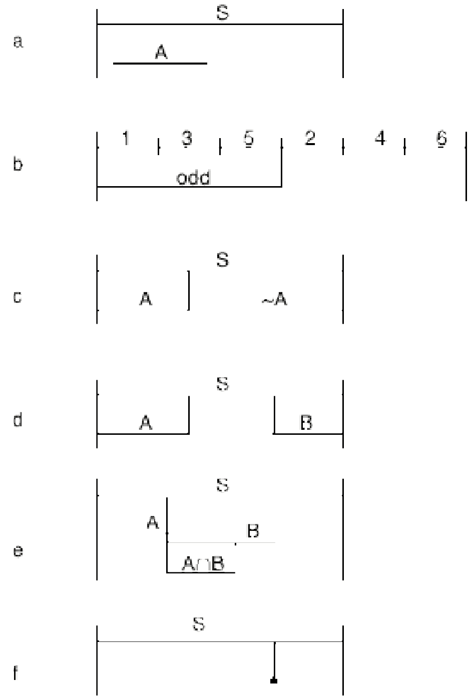

Figure 9. Set theory interpretation of PS diagrams.

Set theory

In PS diagrams parallel labeled horizontal line segments ('lines' for short) represent sets. Figure 9a shows a set, A, in relation to the universal set (or space), S. The common horizontal space occupied by A and S represents the elements that both sets contain (imagine a vertical column with sides intersecting the ends of A). As S is the universal set no parts of A can lie beyond the ends of S. The faint vertical lines are sometimes used to show how particular sets are projected to or from each other, Figure 9. Short sections are labeled to explicitly denote particular elements, such as Figure 9b, which identifies odd numbers from the first six integers. As S is the universal set, the complement of A is simply that part of S that does not project into A, so Figure 9c shows that $A \cup \sim A=S$. Figures 9d and 9e show disjoint and intersecting sets. The intersection of A and B, $A \cap B$, is explicitly shown in Figure 9e. Null or empty sets are represented by lines of zero length, namely points, as shown in Figure 9f. Although of zero length, the horizontal location of a point may be meaningful in the context of the other lines for other sets.

Other properties of sets can be shown with similar diagrams. Figures 10 and 11 are diagrams for commutativity of sets, Equation 1, and de Morgan's laws, Equation 2, respectively. The diagrammatic counterpart of the equality of the symbolic expressions are the two halves of the diagrams, top and bottom, that share a common "result" line (thick line in the middle).

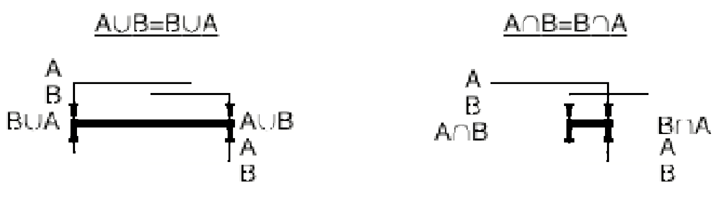

Figure 10. Commutativity of sets.

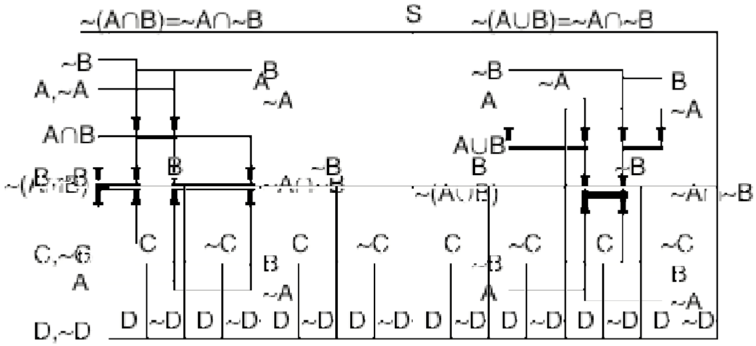

Figure 11. de Morgan's Laws.
Figure 13. PS diagram for all possible intersections of four sets.

Similar diagrams can be drawn for associativity, idempotence, distributivity, the laws of excluded middle and contradiction, identity, absorption, involution, equivalence and symmetrical difference.

Arrows are included in PS diagrams to aid interpretation when the lines are not arranged following the top to bottom convention or when particular elements are to be highlighted.

In relation to set theory, PS diagrams are in effect one-dimensional Venn diagrams. The PS diagram is drawn by partitioning each set into two nested subsets for the other main sets in S and their complements. In Figure 12 there is a line for B and ~B under each of the lines for A and ~A. The selected subsets with thick lines in Figure 12 map onto the equivalent (numbered) subsets of the Venn in Figure 1. As the order of elements in a set is arbitrary, the PS diagram may be re-drawn with all the lines for C and its complement grouped together side by side at the top, with A and ~A broken into eight sections at the bottom, for instance. If the bottom layer is ignored or not drawn, then the PS diagram in Figure 12 represents S containing only subsets A and B, which happen to intersect. The intersection is the left quarter of the PS diagram.

This division of lines into several parts for subsets makes the PS more cumbersome to construct than Venn diagrams, in the first place, but there may be advantages in the interpretation of the diagram. For example, try locating the line/area standing for ~A∩B∩~C∩D in the PS diagram in Figure 13 and the Venn diagram in Figure 2. These are special cases as every set intersects every other set. To depict situations where some but not other sets intersect with the PS diagram it is a matter of deleting columns for those conjunctions that are not applicable. With the Venn diagram it is necessary to reposition the circles with respect to each other or even to change the shapes representing each set.

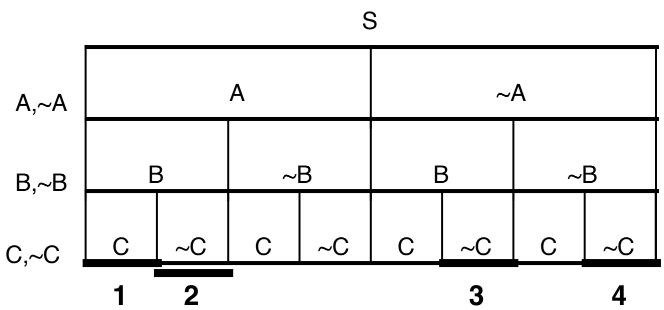

Figure 12. PS diagram for three intersecting sets.

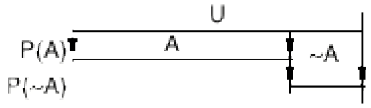

Figure 16. Sum of complementary probabilities

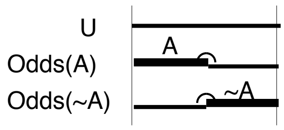

Figure 17. Odds interpretation of PS diagrams.

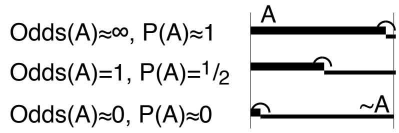

Figure 18. Extreme probabilities as odds.

Probability, Relative Frequencies and Odds

Information about the quantities of objects in sets, or the relative frequency of elements, or the probability of events are encoded in PS diagrams by interpreting the lengths of lines as magnitudes. Figure 14 is a basic PS diagram with lines U and A. Under a quantity interpretation the length of line U may, for instance, represent a total of 1000 items and the length of A represents 750 items of type A. The ratio of the length of A over U gives the relative frequency of A, $rf_u(A)=750/1000$.

Under a probability interpretation, U is taken to be of unit length and the probability of A is given by the ratio of the length of A to U, P(A)=0.75. This constraint captures the second axiom of probability theory, Equation 4. The arrows from U to A indicate that it is the ratio of A to U that is of interest.

Figure 15 shows the extreme cases when the length of A is equal to U and when A is a point, which represents the probabilities of A being equal to unity and zero. This illustrates the range of probability values as given in the first axiom of probability theory, Equation 3.

Figure 16 shows that the probability of A and ~A equals unity, Equation 6.

PS diagrams provide an interpretation of odds, as shown in Figure 17. The overall length of the A and ~A lines must equal the length of U and the small semicircle is used to indicate that an odds ratio is being considered, the ratio of the length of A to ~A. Figure 17 also shows the odds against A, odds(~A), where the ratio is the length of ~A to A. Taking together both the probabilistic and odds interpretations in Figure 18, the diagrams encode the relation between the probabilities and odds, as given by Equations 13 and 14.

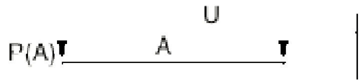

Figure 14. Basic PS Diagram

Dependent Events: Joint and Disjoint

The first of the major classes of probability situations (Figure 3), to be considered, are those with dependent events, Figure 19. PS diagrams use different methods to encode dependent versus independent situations. In PS diagrams for dependent situations the relations among events are encoded by the horizontal relations between the lines for individual probabilities. The diagrams take into account the relative horizontal position of the lines and probabilities are computed by operations that, in effect, add or subtract lengths of line segments from each other.

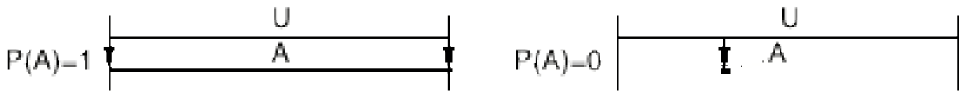

Figure 15. PS diagrams for extremes of probability

In Figure 19, events A and B are joint, and in Figure 20 they are mutually exclusive. The same events A and B are considered in each group of diagrams and for easy reference their lines are shown at the top of each diagram, Figure 19i and 20i. The arrows showing projections of A and B from U have been omitted for clarity and, in general, when a line stands alone in this manner its probability is to be taken with respect to the Unit line, by convention. To find the probabilities of particular events the relevant sections of the lines for A and B are selected, following the set theoretical interpretation above.

Joint and disjoint events are represented in PS diagrams by overlapping or non-overlapping lines, respectively. The magnitude of P(A∩B) is the ratio of the length of the line for A∩B divided by the Unit line. In the case where events are at not disjoint there is an overlap of the lines, Figure 19ii. For disjoint events, however, there is no overlap, so A∩B is empty and the probability is zero, Figure 20ii. Although the lines may not overlap, they may be drawn so that they abut, for the sake of convenience. The probability of A∪B is then the ratio of the length of line between the “free” ends of lines A and B to the length of U, Figures 19iii and 20iii. This encodes Equation 5, for the third axiom of probability theory. Figures 19iv and 19v, shows diagrams for the probabilities P(B|A) and P(A|B) for joint events, and similarly for Figure 20iv and 20v, with disjoint events.

Combining the diagrams for P(A∩B) and P(A∪B), in Figure 19, they encode the relation given in Equation 7, as shown in Figure 21. The probability for events that are not disjoint is less than the sum of the individual probabilities, because of the overlap of the lines for A and B, which represents the intersection of the events. Thus, the joint probability is given by the ratio, to the U line, of the lengths of A plus B minus A∩B. Each term in Equation 7 is represented by a line in Figure 21. Further, why Equation 7 should reduce to Equation 5 in the switch from joint to disjoint events is demonstrated by changing the diagram in Figure 21 so that A and B merely abut, so there is no longer an overlap and the term in the equation for P(A∩B) is a point (zero length).

Multiplication rule and the compound ratio rule

The overlap between two events underpins the representation of conditional probabilities. The value of a conditional probability is given by the ratio of the length of the lines for the overlap divided by the length of the line for the 'given' event. Figure 19iv/v shows examples for P(A|B) and P(B|A). Note that the base line (denominator) is not the Unit line in this case. Using the same logic, prior probabilities are also really conditional probabilities but with the Unit line as the 'given' event: P(A)=P(A|U). This fact is useful for understanding how a PS diagram encodes the multiplication rule, Equation 8a/b.

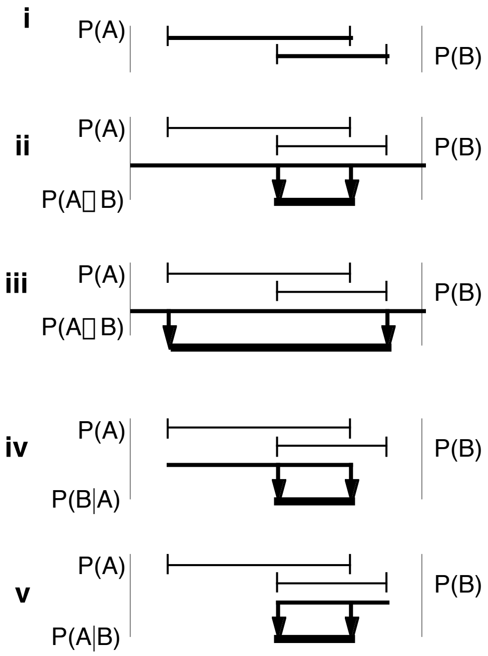

Figure 19. Dependent joint events

First, a simple geometric relation needs to be introduced. Consider in Figure 22 the relation among the ratios of lengths N:M, N:X and X:M. N:M can be considered as composed from N:X and X:N, because N/M=(N/M)(X/X)=(N/X)/(X/M). This will be called the compound ratio rule.

Now consider Figure 23, which depicts the multiplication rule for conditional probabilities. A and B intersect, so the probability of the conjunction is given by the ratio of the thick line in the middle for A∩B to the line for U. The conditional probability of A given B is represented by the ratio of that part of the line for A that overlaps B divided by the length of the line for B. As P(B) is the ratio of the lines for B and the U line, the compound ratio rule is applicable, with the mappings of the lines for events to the lines in compound ratio rule diagram shown on the right of Figure 22. Thus the diagram encodes the multiplication rule, Equation 8a (or 8b), by inserting an extra line for P(A) (or P(B)) into the diagram for the conditional probability of the conjunction.

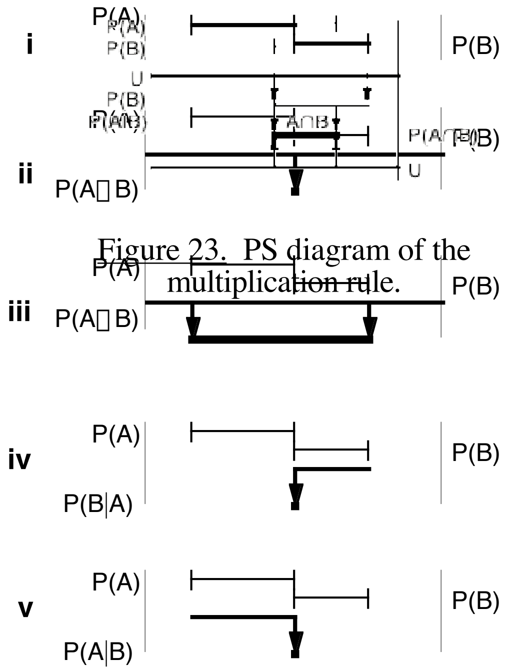

Figure 23. PS diagram of the multiplication rule.

Figure 20. Dependent disjoint events.

Independent Unlinked Events

PS diagrams for dissimilar-unlinked independent events are shown in Figure 24. To encode one trial following another, or two trials occurring separately, the PS diagrams for independent events: (i) have separate lines for each trial; (ii) include the possible outcomes of each trial; (iii) show how each outcome of the first trial affects the overall probability of the outcomes of the second trial but not the individual probability of the events within the second trial. All this is done by scaling the lines for the events of interest in the second trial with respect to the lines for the outcomes of interest in the first trial.

Figure 24 shows various unlinked independent probability relations for events A and B (Figure 24i). Consider the diagram for P(A&B), Figure 24ii. The line for the outcome A in the first trial is drawn. This line is then taken as the base, or 'unit', line for re-drawing the overall line for the second trial, with events B and ~B. Thus the probability P(A&B) is the ratio of the length of the line for A&B to the length of the unit line (which is not shown, but is given by the vertical bounding lines of the probability space). This scaling operation is a geometric form of multiplication, so the diagram encodes Equation 9.

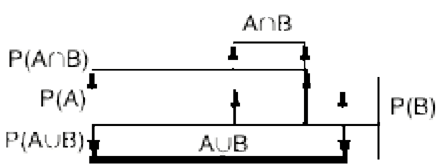

Figure 21. Combining diagrams for. P(A∩B) and P(A∪B)

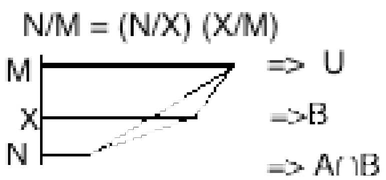

Figure 22. Compound ratio rule.

The diagram for P(AvB), Figure 24iii, uses the same scaling principle but is a little more complex example. Again the line for the second trial, including B and ~B, is scaled to the line for A in the first trial. The line for B and ~B is also scaled to the line for ~A. Thus, the line for AvB is given by the overall length of all the segments for which either A, B, or both, are true. From the left, this is composed of segments for A&B, A&~B and ~A&B. The probability P(AvB) is the ratio of the length of this composite (thick) line to the Unit line. The diagram encodes Equation 10. Further, the diagram also encodes Equation 12, which relates AvB to not ~A&~ B, because the line for AvB covers all of the unit line except the segment for ~A&~B.

PS diagrams for conditional probabilities of unlinked independent events are constructed in a similar fashion, with the scaling of the lines for one trial to the lines for events of interest in the other trial. However, the comparison of the lengths of the lines is not to the unit line, but to the line that is the basis of the conditional, as shown in Figure 24iv/v. Thus, the magnitude of the probability of P(A|B) equals P(A) (and P(B|A)=P(B)), because of the scaling has reduced the length of the line for A in direct proportion to the line of B. Thus Figure 24iv(v) encodes Equation 11a (and 11b). In this case, the resulting magnitude of the probabilities is unchanged, but the scaling procedure is essential to the PS diagram because it captures the underlying process that generates the probabilities.

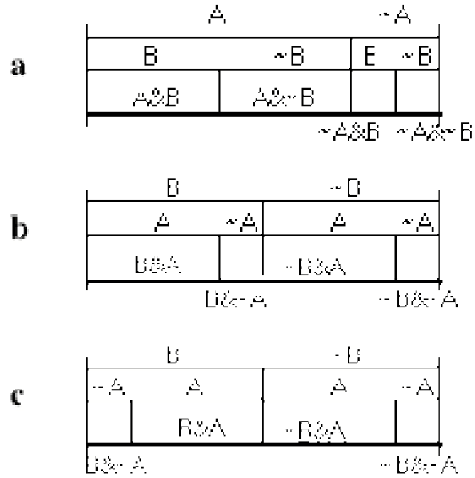

Figure 25. Equivalent PS diagrams for independent unlinked events.

The order of construction with respect to the trials does not affect the underlying structure of the diagrams for events. Compare Figures 25a and 25b in which the trials for A&~A and B&~B are reversed. The same four overall outcomes are obtained and they have equal length lines and probabilities. The scaling of the lines for the second trial with respect to the first has not affected the independent probability of any outcome of the second trial. The line for A in Figure 25b, for instance, has merely been split into two segments but with a constant overall length, equal to that in Figure 25a. Further, rearranging the diagram slightly in Figure 25c, so that the two parts of the lines of A are side by side, it becomes clear that P(AvB) is equal to the overall lengths of P(B) and P(A) (end to end), less the length of P(B&A), because A and B overlap. This encodes Equation 10.

For similar-unlinked trials, the lines for each trial are repeated on successive rows, for example Figures 26a, 28a and 29. The lines are scaled as for all PS diagrams for independent situations and the number of

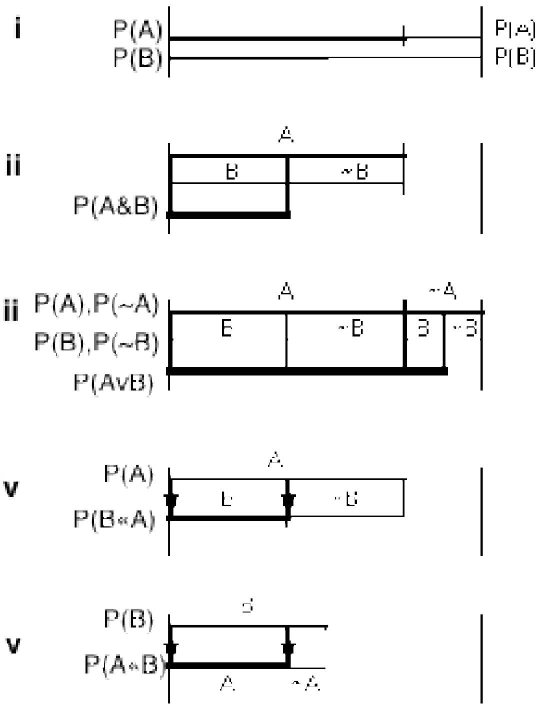

Figure 24. Independent Unlinked Events

repetitions of the lines on the second trial is the same as the number of possible outcomes. In Figure 29, the A-B-C line is repeated three times on the second trial because there are three different events.

Linked Independent Events

This class of situations occurs when particular outcomes of one trial determines the possible events that may occur in the next trial, but the chances of the events within the second trial are not influenced by the first trial. Selection of items from a group without replacement is one form of linking, say beads from a bag. On the first trial there is a choice of n beads with equal probabilities P(U)/n=1/n of selecting any one at random. The second trial is linked to the first as there are now n-1 beads. The probability of choosing any bead in the second trial is 1/(n-1), but the particular bead or its chance of being selected is not a function of which bead was chosen in the first trial.

The encoding for linked independent events uses the same scaling procedure as unlinked events. To begin, Figure 26a shows the PS diagram for the similar-unlinked independent case of the double selection of three beads (A, B, C) with replacement between selections. The line for A-B-C is simply re-scaled and repeated three times on the line for the second trial under each of the three possible outcomes of the first trial. However, for linked events the choice of lines to be scaled is contingent upon the nature of the linking. Figure 26b, shows the PS diagram for a set-linked case when there is no replacement in the selection of beads. As one bead has been removed (e.g., A) its line segment is excluded. The remaining line (B-C) is then re-scaled and located on the line for the second trial under the line for the bead that was removed in the first trial (A). The diagram is completed in the same way for the other beads.

Another example of set-linking is shown in Figure 26c, which is the example given in Figure 3. A die is thrown if a tail comes up on the toss of a coin, otherwise the coin is tossed again. Figure 26d is the example from Figure 3 of the probability-linked case, where coins that are biased towards heads or tails are tossed if the initial toss of a fair coin is a tail or a head, respectively. An example for the combined case of both set and probability linked independent events is shown in Figure 26e. In the second trial, the chances of a '6' or a head are disproportionately high compared to the other numbers or to a tail.

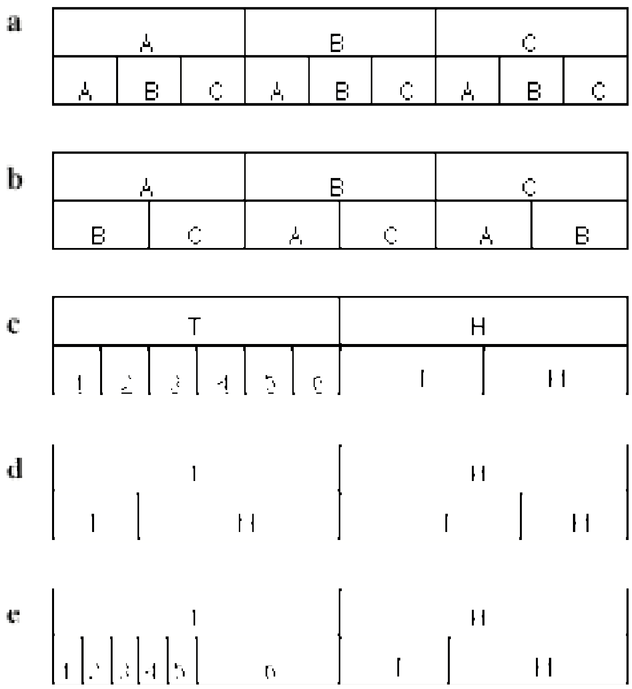

Figure 26. One unlinked and four linked independent events.

The structure of the PS diagrams, the way they encode the axioms and propositions of probability theory and how they are used to model probabilistic situations, have been considered in detail to give an impression of the validity and completeness of the system. Introducing students to PS diagrams would not initially give such detailed explanations of the nature of the geometric constraints and their interrelations. Figure 3 summarizes the different methods of encoding the laws of probability in PS diagrams for the different classes of situations.

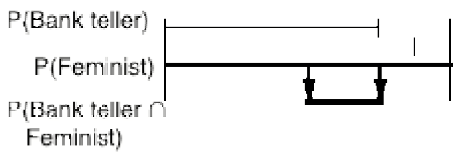

Figure 27. PS diagram for "feminist bank tellers".

MODELLING WITH PS DIAGRAMS

How PS diagrams are used for problem solving and their potential for learning are considered in this section, by applying the diagrams to a range of counterintuitive situations and hard problems. The chosen examples are hard for naive reasoners and for students who have studied probability (Kahneman, Slovic, and Tversky, 1982; Shaughnessy, 1992). As claims are being made about the general nature and scope of PS diagrams a wide range of examples are considered to provide a good sampling of the potential application of the diagrams.

Features of Some Probabilistic Situations

Because PS diagrams encode the laws of probability using the internal structure of diagrams, such that each diagram represents a particular case, they often allow the features of situations to be explored through alternative interpretations of the diagram or by contrasting diagrams. Some examples of this use of PS diagram, are considered.

Conjunctions

Given appropriate contextual traps, naive reasoners will sometimes claim that the probability of a conjunction is greater than the probability of one their parent stem events (Kahneman, Slovic, and Tversky, 1982) . For example, primed with the information that a woman is "bright, single, 31 years old, outspoken and concerned with social justice" college students often rate the statement "She is a bank teller and feminist" as more likely than "She is a bank teller". It is predicted that showing or having reasoners construct the PS diagram for conjunctions of dependent events may better help them avoid such conjunction fallacies than knowing and using Equations 7 and 8. The equations do not relate the probability of the conjunction directly to the probability of the stem events, but require the consideration of either the probability of the disjunction or a conditional probability. However, in Figure 27, which models this situation, the overlap of the lines for the stem events gives the probability of the conjunction. It is clear that the probability of the conjunction can never be larger than the line for the smallest of the stem events. Further, the greatest probability occurs when one event is completely subsumed by the other, so that the overlap equals the length of one of the events, say, when all feminists are bank tellers.

Combinations and permutations in repeated selections

Consider situations in which there is repeated selections of objects. Aspects of such cases include: (i) combinations versus permutations of objects; (ii) numbers of objects, n, and numbers of selections, k; (iii) numbers of particular combinations or permutations; (iv) whether there is or is not replacement of objects. How are these different aspects interrelated? The formulas for independent events, Equations 9 to 12, and formulas for combinations and permutations, Table 1, can be used to begin to explore the relations among these things, but with difficulty.

An approach with PS diagrams is to draw and interpret diagrams for particular cases and to attempt to generalize any patterns found bearing in mind the constraints of the diagrams. Figures 28a and 28b show the PS diagrams for selections with repetitions and without, respectively, for successive values of k for the case of n=4. The comparison of diagrams reveals some interesting aspects of the two situations.

The number of permutations of events increases rapidly in both cases, as shown by the number of distinct columns on successive rows. In Figure 28a there is a geometric increase, which is more rapid than the factorial increase in Figure 28b. As the number of trials grows the chances of a given permutation decrease in inverse proportion to the number of permutations, which keeps the overall probability of the whole space at unity. Only three trials are shown in Figure 28a, because of the physical limits of drawing distinguishable segments given the geometric decrease in the size. All four trials can be shown for the case without repetition in Figure 28b.

The shading in Figure 28a picks out one set of unique combinations, starting with combinations on the left of the diagram and skipping over any columns that have combinations which are already highlighted. The number of combinations is clearly increasing, but the rate is much less than the geometric increase for permutations. Notice how the pattern of shading on each successive row is generated from the previous row in a fractal-like manner. Such patterns, once spotted can be used as a method to enumerate the set of combinations for an arbitrary number of trials. Figure 29 shows the unlinked independent toss of up to six coins, showing a similar fractal pattern. Notice how just one new combination is added on each trial, which mirrors the addition of just one new coefficient on each successive row of Pascal's triangle.

In contrast, in Figure 28b for no repetitions, on successive selections the number of combinations initially increases from unity, to n, reaching a maximum, then dropping back to n

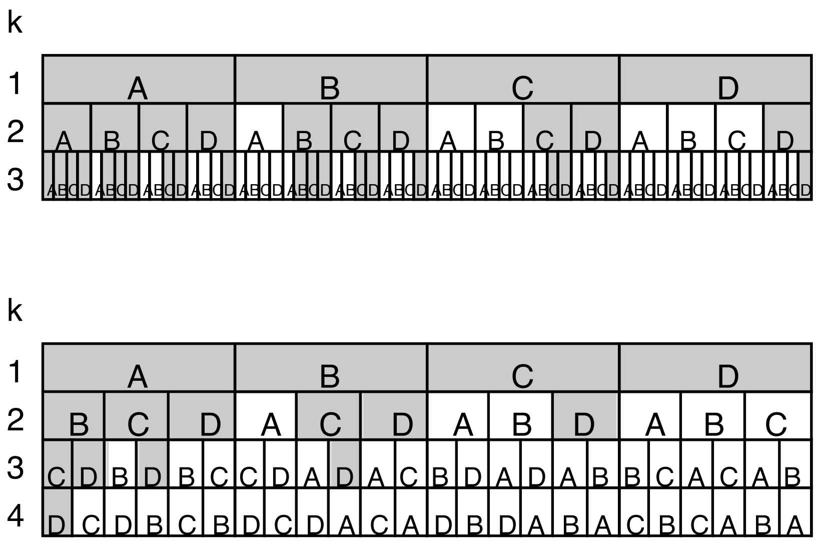

Figure 28. Combinations and permutations in PS diagrams.

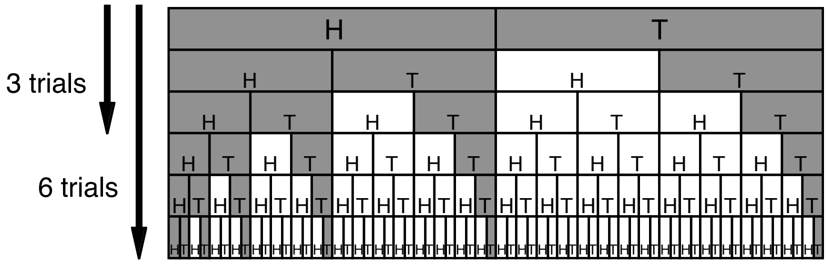

Figure 29. Three or six toss of a coin.

and finally unity (corresponding to 4!/(4!0!), 4!/(3!1!), 4!/(2!2!),4!/(1!3!) and 4!/(0!4!)). As there are only four objects, by the fourth selection all have been picked, so each of the 24 (permutation) columns in Figure 28b must logically be the combination ABCD. The logic of the process of selection without replacement means that the number of permutations for k=n-1 selections is equal to that of permutations for k=n selections, because there is no choice for the final selection. This is shown by the lack of subdivision of the line segments between the penultimate and ultimate lines in Figure 27b (k=3 and k=4).

On a given trial, the probability of any given permutation is less than the case where there is no replacement than for when there is replacement. By definition, the probability of the selection of particular objects when there is independence and no linking is the same from trial to trial, which is shown by the constant total length of the line for each object, despite the lines being subdivided, Figure 28a. Perhaps somewhat surprisingly, the same is true in the linked case in Figure 28b, despite the fact that in different branches down the diagram the earlier selections affect the possible choices of later selections. Overall, the regularity that holds horizontally across the diagrams means that the local differences in particular branches are evened out for the PS diagram as a whole.

Sampling

Sample size affects the probability of permutations and combinations, although naive reasoners often do not believe so (Shaughnessy, 1992). For example, the probability of obtaining 3 heads in 3 tosses of a coin is not the same as the probability of six heads in six tosses of a coin. Figure 29 illustrates this using a PS diagram, with the width of the column for the combination with all heads, to the left, decreasing with each new trial in the sample.

As each column in Figure 29 stands for a unique permutation, the diagram shows that the chances of obtaining HHHHHH in order is no more or less likely than HTHHTT, even though the latter permutation may appear as more representative to naive reasoners (Shaughnessy, 1992). Similarly, the fallacy of negative and positive recency effects is demonstrated by the equal division of each line segment on each trial, irrespective of the number of preceding trials. This is required by the scaling rule for independent trials, which is applied locally to each line segment in the diagram. However, on a global scale the uniform application of the rule means that the chance of a particular permutation, such as successive heads, diminishes geometrically with the number of trials.

Figure 29 may be used as an introduction to the Binomial distribution. After two trials there is one set of HH, two of HT and one of TT. After three trials the sets are: 1 HHH, 3 HHT, 3 HTT , 1 TTT. By collecting series of outcomes in this manner the coefficients of the terms in the binomial expansion can be found. As the widths of the lines for unique events are equal on a particular trial, the total width and hence probability of a given combination of events can be

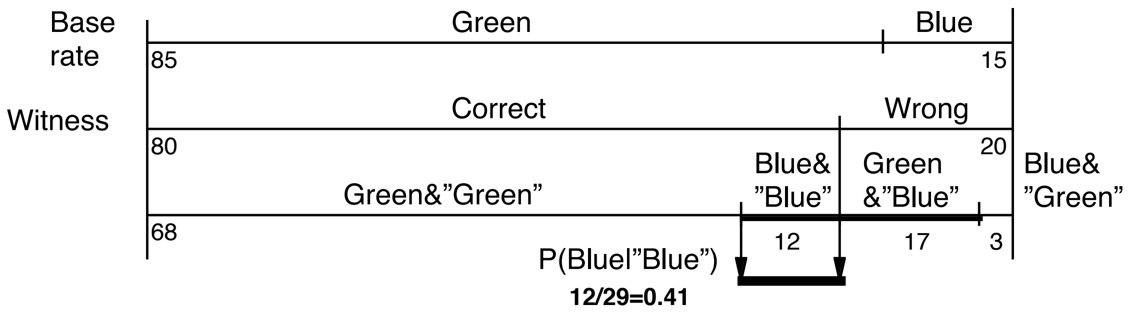

Figure 30. PS Diagram solution to the Blue cab/Green Cab problem.

found. Thus, the PS diagram provides an illustration of the relation between the probabilities of the possible outcomes and the form of the binomial distribution over different numbers of trials.

Hard Problems

PS diagram solutions to some classic problems in probability are presented to illustrate how the diagrams can be used to model complex situations.

Testing Situations

The Cab problem gives many reasoners great difficulties. As the situation is independent and unlinked, the PS diagrams to model this case will have the same general form as Figure 25 but lengths of the lines will stand for the given percentages, as shown in Figure 30. Using the scaling rule the base rate line is scaled to the lines for the correct and incorrect identifications of the witness. There are four possibilities, two for correctly identified cabs and two for incorrectly identified cabs, with their respective probabilities (%) shown in Figure 30. The answer is the conditional probability that the witness was correct when saying "blue", P(Bluel"Blue"), which is the ratio of the line for Blue&"Blue" (correct identification) to the sum of the lengths of the lines where "Blue" was said irrespective of the truth of the identification (i.e., 12/(12+17)). The PS diagram configuration for the conditional is shown in Figure 30 by the medium lines and the thick line projecting from it (cf. Figure 3iv/v).

The PS diagram for this problem has an advantage over the tree diagram, Figure 6, or the contingency table, Figure 7, because it shows more clearly what calculation to do once the intermediate probabilities (four products) have been computed. The diagram reveals the implicit conditional outcome that is hidden in the problem statement and makes it clear that the base rate affects the chances of correct and incorrect identifications. One might deliberately promote this PS diagram configuration for independent unlinked events as a general model for test situations, where there are possible false positives and false negatives, in addition to correct test results. By varying the base rate and the identification percentage students can explore the effects of the reliability of alternative tests in different circumstances.

There is a close similarity between this PS diagram and the “bar” representation for this specific type of problem used by one of participants in Gigerenzer and Hoffrage’s (1995) study. Such visual representations appear to be used by some reasoners when probability problems are presented in frequentist with an associated improvement in their quality of making Bayesian inferences. This suggests, indirectly, that PS diagrams may also improve reasoning on problems about such “testing” situations.

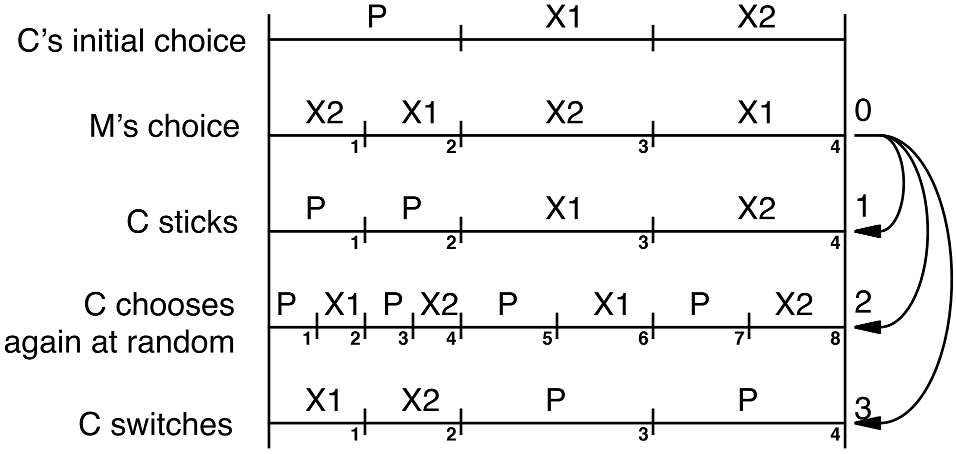

Figure 31. The probability space for Monty’s dilemma

Complex Linking in Monty's Dilemma

Naive reasoners doing this problem often think that when Monty opens his door the probability immediately increases from 1/3 to 1/2 and that sticking with the original door or switching will have no further effect on the probability of winning. Shaughnessy (1992) claims "this problem highlights the difficulty with what is actually known, when it is known, and how the new information obtained is used" (p. 475). The solution with PS diagrams puts a different perspective on why the problem is hard.

This problem is a fine example of the complex linking of events across trials. The linking occurs because (i) Monty makes a choice that is influenced by the contestant’s first choice and (ii) contingent upon Monty's selection the contestant has three options: (1) stick with the original door, (2) flip a coin to choose between the two remaining closed doors, (3) switch to the other door. The PS diagram solution is shown in Figure 31. The three lines on the top row are the contestant's (C's) possible choices. The lines are labelled P, X1 and X2 for the prize and the two dummies. The second row shows Monty's, M's, choice contingent on C's choice. If C initially picks the door with the prize behind it, P, M can select either of the other doors, $X1_{0-2}$ or $X2_{0-1}$ (subscripts are of the form ‘n-m’ for the m^{th} possibility of contestants n^{th} option, as shown in Figure 31). Alternatively, if C initially picks a dummy door (e.g., X1) then M can only pick the other dummy door (i.e., $X2_{0-3}$).

Option 1: The third row shows the consequences of sticking to the original choice. The probability of winning is still 1/3, not 1/2, even though C knows that X1 or X2 is a dummy. There are two ways to explain this, either using global or local interpretations of this PS diagram. First, in terms of the whole probability space, M’s opening of a dummy door does not eliminate any of the possible outcomes overall, because of the particular way the outcomes are linked in a cross-wise fashion over the first three rows of the diagram. P remains the same proportion of the whole space. For instance, M choosing $X2_{0-3}$, given C initially picked X1, does not mean that C might not have initially picked X2 (and hence forced M to choose $X1_{0-4}$). Second, consider the local conditional probability of C picking P given that M opened X1. This is given by the ratio of the length of P_{1-2} to $X1_{0-2}$ plus $X1_{0-4}$, which is 1/3. Similarly, with P given X2 ($P_{1-1}/(X1_{0-1}+X1_{0-3})=1/3$). As both cases are 1/3 then the chances of P must also be 1/3.

Option 2: The fourth row shows the possible outcomes of randomly re-selecting one of the closed doors, after M's choice. The chances of winning goes up to 1/2. Again this may be explained by global or local interpretations of the diagram. First, to globally determine the chances of C getting P, all the outcomes are enumerated for each of M's choices. For example, if C initially picks P, then M can randomly choose either (a) $X2_{0-1}$ or (b) $X1_{0-2}$, which leaves C with the choice of (a) P_{2-1} or $X1_{2-2}$ or (b) P_{2-3} or $X2_{2-4}$, respectively. The sum of the length of all the resulting P line segments ($P_{2-1,\ 2-3,\ 2-5,\ 2-7}$) is one half the Unit line. Alternatively, locally examining the options C has after each of M's choices, there is always an equal choice between two events with one of them being P. Hence, the probability of C winning is 1/2 as every set of outcomes has a probability of 1/2.

Option 3: When C switches the probability of winning goes up to 2/3. Yet again, this may be explained globally or locally. First, consider each of C's options in response to M's choices. If C initially picks P and M picks either $X2_{0-1}$ or $X1_{0-2}$, this means P will switch to $X1_{3-1}$ or $X2_{3-2}$, respectively. If C initially picks either X1 or X2, M picks $X2_{0-3}$ or $X1_{0-4}$, respectively, so on switching, C will pick $P_{3-3/3-4}$ in both cases. Thus, two thirds of the probability space is occupied by P on the fifth row. Alternatively, the conditional probability of C having P given M selecting X1 or X2, are individually 2/3, hence overall the probability must be 2/3.

The examination of the problem space provides an explanation why this case is so hard to comprehend. The complex linking of events between trials, with some involving a forced choice and others a random selection with the application of the scaling rule, makes the task of following what is happening extremely difficult, because there are so many events and contingencies to bear in mind. Thus, even experts in the domain who attempt linear verbal reasoning about the problem usually fail to cover all of the possible sequences of events. Such reasoning relies on the local consideration of particular sequences, but it is essential to consider all the paths through the probability space in order to determine the overall probabilities of the different outcomes.

Three Prisoners and Bayes' Theorem

Shimojo and Ichikawa (1989) contrast the correct Bayes' theorem solution to the 'problem of the three prisoners' with the subjective theorems that they found that people often apply when attempting solutions. PS diagrams provide an alternative to Bayes' theorem that makes the solution less counter-intuitive.

> <u>Problem of the Three Prisoners.</u> The people, A, B, and C were in Jail. A knew that one of them was to be set free and the other two were to be executed. But A did not know who was 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 name the person, B or C, who is to be executed". Accepting this argument after some thinking, the jailer said "B will be executed". Thereupon A felt happier because now either A or C would go free, so A's chance had increased from 1/3 to 1/2. Is A's happiness reasonable?

There are obvious similarities between the structure of this problem and Monty's Dilemma. Consider first the solution using Bayes' theorem. This requires the prior probabilities, which are assumed to be equal,

$$P(A) = P(B) = P(C) = 1/3 \,. \qquad 15$$

Also, the conditional probabilities of the jailer saying "B" given the different options of the person who might be released are,

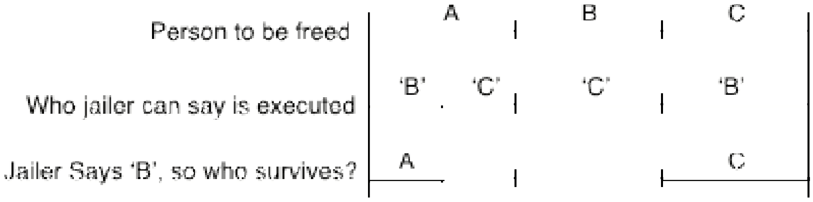

Figure 32. PS diagram for the Problem of the Three Prisoner

$$P(\text{"B"}\text{«}A) = 1/2\ , \tag{16a}$$

$$P(\text{"B"}\text{«}B) = 0\ , \tag{16b}$$

$$P(\text{"B"}\text{«}C) = 1\ . \tag{16c}$$

If A is to be freed the jailer has a 50-50 chance of whether to say "B" or "C". If C is to be freed the jailer has one choice and if B is to be freed there is no choice. To compute P(Al"B") this version of Bayes' theorem is applied,

$$P(A\text{«}\text{"B"}) = \frac{P(\text{"B"}\text{«}A)\ P(A)}{P(\text{"B"}\text{«}A)\ P(A) + P(\text{"B"}\text{«}B)\ P(B) + P(\text{"B"}\text{«}C)\ P(C)}\ . \tag{17}$$

Substituting the values from Equations 15 and 16a-c yields a probability of 1/3.

One approach to the solution to the problem with PS diagrams is to enumerate all six possible sequences of events of who might be spared or executed and then to eliminate those sequences that are false, because it is known that B will be executed. Then conditional probabilities among the remaining outcomes can be considered. This general strategy was used in Figure 31 for Monty's dilemma. A simpler alternative is to consider the outcomes and work backwards. Figure 32 shows the PS diagram solution to the problem. The first row shows the possible outcomes, in terms of who will be free. The second row shows who the jailer can say will be executed, given that he will not say it is A. When the jailer says "B" there are thus two possibilities of who else will be executed as shown by the two thick lines in the third row. Thus, the chances of A being freed are not greater even when it is known that B will be executed.

The problem can be made even more counter-intuitive by changing the prior probabilities of each person being free such that P(A)=1/4, P(B)=1/4 and P(C)=1/2 (Shimojo and Ichikawa, 1989). In this case the chances of A being freed drop to 1/5, surprisingly. The application of Bayes' theorem does not easily explain why. Figure 33 is a re-drawn PS diagram for this situation, showing how the different prior probabilities alter the probabilities lower in the diagram. Note that the lines B and C are equal under A, because it is assumed that who the jailer will say is to be executed is random and unrelated to the relative prior probabilities of the B and C. If for some reason it was in proportion to the prior probabilities, then the application

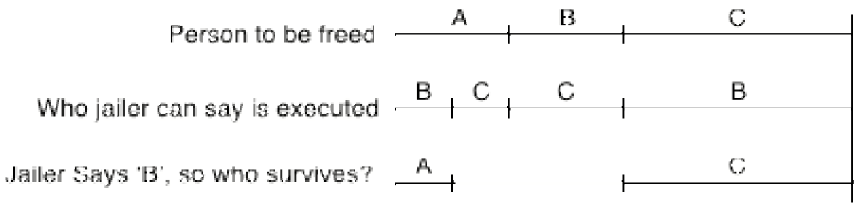

Figure 33. PS diagram for the Three Prisoner Problem with unequal prior probabilities

of the scaling rule would make the chances for A even worse.

This example also demonstrates that PS diagrams may be used to counter the Falk Phenomenon (as described by Shaughnessy, 1992) — the reluctance of naive reasoners to entertain the possibility of conditional probabilities in which the conditionalizing datum, the given, occurs later in time than the item of interest. Although it is sometimes useful to arrange successive rows vertically in the same order as the temporal occurrence of the trials in the cover story, this is not essential in PS diagrams.

ALGEBRA VERSUS PS DIAGRAMS

PS diagrams have been introduced and a range of examples considered to demonstrate the scope and coherence of the system as a representation for probability theory. This section makes predictions about the potential benefits of PS diagrams for problem solving and learning, in comparison to the traditional algebraic approach, on the basis of each of the six characteristics of effective representations. These psychological predictions about the representations are made by considering the formal structure of the notation they employed in relation to the general nature and constraints of human reasoning and learning. Different cognitive implications flow from the formal differences in the representations.

G1 Integrating Levels of Abstraction

The levels of abstraction are not well integrated in the traditional algebraic approach, because information at different levels of abstraction are distributed over separate expressions. Sets of equations are used to encode the abstract laws and others used to assign values to particular variables. As in any complex domain, it is common for learners to lose sight of the concrete interpretation of what is happening whilst they are doing the manipulation of formulas needed to meld the laws and values. For example, in the Cab problem there is no simple explanation as to how the base rate information interacts with the witness's accuracy. The algebraic formulas for independent events do not make this evident, because there are many combinations of events to consider, which is why contingency tables (Figure 7) or tree diagrams (Figure 6) are used to help co-ordinate the information. Bayes' theorem is an alternative method, but it too is semantically opaque, because formulas such as Equation 17 do a good job of obscuring the underlying structure of the problem and hence effectively separating the levels of abstraction.

In PS diagrams information about the abstract laws and values for particular cases are incorporated into the same expressions. The laws are encoded in a diagram by geometric and spatial constraints, such as the overlapping lines for joint sets and the scaling rule for independent events. The elements in the same diagram also represented magnitudes as the lengths of lines or by the simple ratio of lengths. The comparison of Figures 32 and 33, shows how, in the complex Three Prisoners problem changing certain values (base rates) affects the final outcomes, but within the context of the overall structure of the situation as determined by the laws of probability (scaling rule for independent situations and linking of contingent events). Similarly, Figure 30 for the Cab problem shows the solution for particular values of a situation, but the overall structure of this diagram stands as a general model for problems involving the testing of hypotheses when false positives and negatives are possible.

PS diagrams integrate levels of abstraction better than the algebraic approach. Thus, it is predicted that users of PS diagrams will find it easier to derive implications of the laws for particular situations and to generate explanations of specific cases based on the laws. In turn, the users are more likely to obtain an understanding of the domain that is structured more by the laws than by superficial features of different cases. Given the same amount of instruction, it is anticipated that learners using PS diagrams will have more expert-like categorization of problems whilst the users of the conventional algebraic approach will be more novice-like (cf. Chi, Feltovich and Glaser's (1981) physics experts).

G2 Globally homogenous and locally heterogeneous representation of concepts

Global homogeneity. The algebraic approach generally lacks interpretative schemes that use the structure of its expressions to encode overarching laws or constraints that hold across the whole domain. There are many examples of this. Many fundamental concepts are given as a single equation or formula that have little direct role in the generation of expressions, such as the first axiom of probability, Equation 3. The necessity of measuring all magnitudes of chance against a datum, including prior probabilities against the unit probability of the universal set (S), is a foundational concept that is not acknowledged in the structure of the algebraic expressions. There are different types or styles of equations for the basic laws of probability. Some laws use only elementary mathematical operations (Equations 2-14), whilst others use more complex operations, such as factorials in the formulas for the numbers of permutations or combinations in repeated trials (Table 1). The need for Venn diagrams, tree diagrams, contingency tables and outcome tables adds to the lack of global homogeneity, because they use different representational schemes (e.g., node-link diagram, table) that are each largely arbitrary with respect to the underlying structure of the domain. Patterns that reflect the same general conceptual distinctions will be different in each of the representations, because they have different formats. All these things taken together are likely to obscure the underlying conceptual coherence of the domain from learners and to be the sources for acquiring erroneous conceptual distinctions.

However, the algebraic equations for probability do represent one important concept in a global homogeneous manner. The notion that all measures of chance apply to some well-defined states of affairs, is shown by means of a set theoretic expression in the parentheses of every probability formula. But compared to PS diagrams this is a minor feature.

The global homogeneity of PS diagrams is apparent, even at a superficial level, in the relatively uniform structure of the diagrams and by the lack of supplementary representations. PS diagrams support all the functions of the supplementary representations in the other approach. More importantly, they provide an explicit unified conceptualization of the conceptual underpinnings of the domain in terms of probability space. (1) The association of a magnitude of probability with every event is achieved by each event being represented by a line whose length is in proportion to its probability. (2) The first axiom of probability theory is inherent in the definition of the probability space and is represented by the width of each PS diagram. (3) The concept that every probability is a conditional probability is reflected by all values of probability being ratios of the length of some line to the length of its base line. In the case of prior probabilities, an event is conditional upon the universe of interest, U, that is the width of the probability space. These features of PS diagrams may help learners make connections between related concepts and gain an overall sense of the overall conceptual coherence of the domain.

Local heterogeneity. The traditional algebraic approach substantially lacks a local heterogeneous representation of concepts, because the notational structure of expressions for different concepts are sometimes the same. The expressions for basic set relations all have the same general notational form:

$$N \square op \square M \,, \qquad 18$$

where N and M are sets and op is a symbol for the set relation, '□', '□', or '|'. The expressions for different relations are distinguished just on the meaning attached by convention to the op symbol.

Similarly, the notational structure of the probability formulas for "and" and "or" relations are the same in the case of dependent and independent events (i.e., Equation 8 and 9, and Equations 7 and 10). Their general forms are, respectively:

$$P(S_{and})=P(S_1) \text{ x } P(S_2)\,, \qquad 19$$

and

$$P(S_{or})=P(S_1) + P(S_2) - P(S_{and}), \qquad 20$$

where S_{and} and S_{or} are expression for "and" or "or" relations between events, and S_1 and S_2 are simple or conditional expressions for particular events. With regard to the major dimensions of probabilistic situations, there is nothing in these expressions that acknowledges the difference between dependent and independent situations, unless attention is shifted to the embedded set theory expressions.

Given the similarity in the notational structure of these formulas, learners are likely to find them rather indistinguishable from each other and, importantly, unlikely to find good cues to associate them with different parts of the space of concepts underlying this topic. To illustrate this point, suppose the right hand sides of Equations 5 to 14 were removed from the text above and listed in a random order. It is a challenging task to correctly sort them out and replace them where they belong in the text. Repeating this task with certain probability textbooks is even more daunting as the representations used are even less locally heterogeneous, because the symbols for set relations (□, □, l) are used for independent as well as dependent situations. Here, the introduction of alternative symbols (&, v, «) for independent events is an attempt to make the conventional algebraic formula satisfy the requirement for the locally heterogeneous representation of concepts a little better.

In contrast, PS diagrams have a more heterogeneous representation of concepts locally, because different diagrammatic constraints encode the various dimensions and sub-dimensions of probabilistic situations (Figure 3). For example, dependent events are a single row in a PS diagram and are distinguished from independent events that have multiple rows for successive trials to which the scaling rule has been applied (Figure 24). Disjoint dependent events are distinguished from joint dependent events by the occurrence of overlapping lines (Figures 19 and 20). PS diagrams for unlinked independent situations have lines that are the same as previous lines (similar-unlinked, Figures 26a, 28a, 29) or that are completely different (dissimilar-unlinked, Figures 25, 30). Linked independent situations, however, have lines that share some features of previous lines but differ in other respects, such as the selection of segments (set-linked, Figures 26c/b, 28b) or different relative lengths of segments (probability-linked, Figure 26d) or both (set-and-probability-linked, Figure 26e). Unlike algebraic expressions, there is usually some notational means to locate where in the conceptual space of the domain (Figure 3) a given diagram belongs.

PS diagrams have a globally homogenous but locally heterogeneous representation of concepts, but the reverse is true for the conventional algebraic approach. Thus, it is predicted that users of PS diagrams will more easily identify conceptual connections between related concepts and will more clearly differentiate concepts that are dissimilar. In turn it is expected that they will gain a better comprehension of the overall structure of the conceptual topology of the domain (Figure 3) and will have better recall of expressions that are applicable to given problem situations. It is possible that PS diagrams for different parts of the conceptual space may stand as icons for the relations that hold in different contexts.

G3 Integrating Perspectives

There are three levels or different perspectives to be integrated in any effective representation for probability: ontological, alternative measures, and viewpoint.

Under the traditional approach, at each level, separate representations or expressions are used for each perspective, and additional notational techniques or coordinating expressions are needed to link the alternative perspectives. For the set theory and probability theory separate representations for the two ontologies are used (e.g., Equation 18 versus Equations 19 and 20), and they are coordinated by the embedding of the expressions of set theory within the probability formulas (e.g., S_1, S_{or} in Equation 20). The different interpretations of measures of chance (probability, odds, frequency) has its own set of equations to encode relations and additional equations are also needed to interrelate the perspectives (e.g., Equations 13 and 14). The same is true for different viewpoints within a problem situation. Individual equations capture particular relations, such as Equation 7 for the union of disjoint events. However, to change the viewpoint, say, to consider the union of the negation of the same events, other probability equations and set theory relations must be applied, including the formula for complementary events (Equation 6) and de Morgan's laws (Equation 2a/b).

The multiple sets of representations, and the different techniques needed to coordinate the alternative perspectives at each level, are likely to prevent users reinterpreting problems to resolve impasses by switching perspectives. In turn, the lack of coordinated application of the alternative perspective is likely to mean learners fail to make good conceptual connections between the different ways of conceptualizing the domain.

In PS diagrams all three levels of perspective are well integrated. At the ontological level, set theory and probability theory both use the same diagrammatic means to represent (i) sets and relations among sets and (ii) events and probabilistic relations among events. A single notational system is employed rather than two coordinated but otherwise independent representations. This is possible because different diagrammatic features of the same notation are used to encode information from the two perspectives: relative position for set theory and length for probability theory.

The different measures of chance are also well integrated in PS diagrams. Accessing a particular interpretation requires different readings of the same expression, rather than separate expressions. It is likely that the particular way of reading a PS diagram will focus attention on different geometric relations between lines and take lengths to stand for different types of quantity. Under a relative frequency interpretation a line represents a number of equivalent things and its length is compared to the line for the number of elements in the universal set. Figure 18 provides probability and odds interpretations of the same diagrams. Both measures are ratios that compare the lengths of two lines, but a different line in each diagram is used as the "denominator" for the two measures.

The same is true of different viewpoints within the same diagram. In a PS diagram of a dependent situation, focusing on the overall length of two lines may give the probability of the union of their events, whilst looking at their overlap gives the probability of their intersection, but comparing the overlap to any one of the lines gives the conditional probability — all in the same diagram. Figure 16 encodes the individual probabilities of an event and its complement and also shows that the probabilities sum to unity. The local and global interpretations in each of the options in the PS diagram for Monty's dilemma is another case in point.

PS diagrams provide a better integration of the three perspective levels. Thus, it is predicted that users of the representation will find it easier to switch between perspectives and thus are more likely to use alternative interpretations to constrain the search for problem solutions or to resolve impasses during problem solving. Easy access to the alternative perspectives of probability may mitigate the effects of biases in reasoning that are strongly exhibited under particular ontologies. In turn, it is likely that learners will develop good conceptual connections between perspectives at the same level and within levels of the same perspectives. The triangulation of the alternative ontologies, measures and viewpoints may provide a deeper understanding of the domain.

G4 Malleable Representations

The generative nature of both the traditional algebraic and PS diagram approaches to probability is clear.

However, the representations under the traditional approach are not malleable, some parts being too flexible and others too rigid. On the one hand, the algebraic representation, per se, and the set theory representation are both highly fluid. Given any expression, many new expressions can be generated using the respective operations of the systems. Choosing the most appropriate operation requires knowledge about the nature of the notational system itself, in addition to some understanding of the domain. For the inexperienced user of the representations, the stream of derived expressions can flow in unexpected directions away from the desired goal. On the other hand, the supplementary representations for the traditional approach are examples of rigid representations. For example, Figure 8 shows the outcomes of throwing two fair dice, but it cannot be directly modified to deal with the probability of different outcomes when two biased dice are considered.

Compared to the algebraic approach, PS diagrams constitute a malleable representation rather than a fluid or rigid system. There are three main reasons for this. First, typically one or two expressions need to be considered when reasoning with PS diagrams, rather than the multiple expressions under the algebraic representation. Thus, the combinatorics of manipulating expressions to derive further expressions are less of a concern. For example, in the PS diagram solution to Monty's dilemma, Figure 31, one large expression is produced by gradually adding new lines for successive trials, whereas under an algebraic approach numerous equations are written for every event of every trial and must then be combined using a tree diagram, for instance. Second, each correct instantiation of a PS diagram represents some possible state of affairs under probability theory, which means that constraints determined by the semantics of the problem under consideration are readily imposed. The chances of selecting expressions that will be irrelevant to the gaols of problem solving are likely to be less with PS diagrams than with the algebraic representation, because algebraic equations are often harder to interpret with respect to the specific requirements of a given problem. Third, PS diagrams perform the function of outcome tables like Figure 8 (e.g., Figure 28 and 29) but they are not rigid. PS diagrams can be directly modified to show related outcomes; for example, the changing of the base rate in the problem of the Three Prisoners from Figure 32 to Figure 33, or modifying Figure 26c to deal with a biased die and coin in Figure 26e.

Thus, for people with the same level of knowledge with respect to domain content, it is predicted that the overall size of the space of (psychological) problem states for reasoners using PS diagrams will be, in general, smaller than the space for users of the conventional approach. The problem space for the conventional approach will be larger because it has flexible and rigid components. Flexibility, by definition, means that many expressions can be generated, so it is more likely that a problem solver will visit more states that each correspond to a particular expression. Substantial rigidity means that supplementary representations must be used and as each system of representation introduces a new sub-problem space, the overall size of the search space will be increased. A smaller problem space means the user of PS diagrams may pursue fewer unfruitful searches, may find it easier to recover from impasses and make less syntatic errors. In turn, this means the learner may see more correct instances and will have more time and opportunity to make conceptual connections.

G5 Uniform procedures

The procedures under the traditional algebraic approach to probability are relatively complex both in terms of their lack of uniformity and compactness. PS diagrams, in contrast, have relatively uniform and compact procedures.

One major reason the algebraic approach does not possess uniform procedures is that supplementary representations are employed, which each require their own particular set of procedures. Constructing a Venn diagram, drawing a contingency table and deriving an algebraic equation all use different sets of operations. With regard to the main representation, the procedures are not uniform because many different techniques are needed in the manipulation of algebraic formulas. This is not just an issue for problems dealing with conceptually distinct parts of the domain, but is a concern for problems in similar situations that merely differ in terms of their complexity. For instance, simple set-linked independent events can be computed by simple application of the basic laws of probability (e.g., throwing a die or tossing a coin on the particular outcome of the toss of an initial coin). However, for a problem that is simply more complex (e.g., Three Prisoners problem), but which is also a set-linked independent event problem, Bayes theorem is a more effective method to use.

Given that PS diagrams integrate all the perspectives that are necessary for reasoning about the domain into a single representation, they naturally have a relatively uniform set of procedures for making inferences and doing problem solving. Whether the diagram is being constructed to encode a certain relation among different sets (e.g., Figure 13), to show various odds ratios (Figure 18), or to capture the complex linking between independent events (Figure 26), the procedures always involve the construction of the probability space with lines arranged horizontally and vertically to model the given relations and situation.

There are two general approaches that may be adopted in the construction of PS diagrams: (1) assemble the diagram from component line segments for events; (2) subdividing the probability space to obtain partitions of the universal set. Further, there will be detailed differences between procedures for alternative relations or situations, because they have different structures, by definition. Nevertheless, these procedures all use a small set of spatial and geometric operations, so the variety of procedures is substantially less than the gross differences among the methods needed for the multiple representations under the algebraic approach.

Thus, for problem solvers with the same level of competence, on a broad range of problem types, it is predicted that the variety of procedures that problem solvers using the PS diagrams will exhibit will be smaller than that of users of the conventional approach. With PS diagrams, common sets of procedures are likely to be found across different classes of problems, solution strategies and individuals. In turn, learning may be enhanced, because there will simply be fewer procedures to master to cover the same content of the domain.

G6 Compact Procedures

One reason that the algebraic representation is inevitably not compact is the distribution of problem solving information among many expressions. Much of the processing of the representations is needed merely to marshal the information into canonical forms that are amenable to available methods of computation. For example, constructing the tree diagram or contingency table for the Cab problem are such activities. For the Bayes theorem solution to the Three Prisoners problem, intermediate expressions for the prior probabilities (Equation 15), conditional probabilities that the jailer will say "B" (Equations 16a-c), and Bayes theorem (Equation 17), typically have to be generated. To compute a different outcome events many of the same inferences must made but for different combinations of given events.

In contrast, PS diagrams are relatively compact, because all the information needed for problem solving is usually assembled in one diagram. For example, each translation of the information from the problem statement of the Three Prisoners goes directly into the "final" solution expressions as successive lines of Figure 32 (or 33), building directly on the previous lines. Once drawn, the same diagram may be used to examine alternative outcomes merely by focusing on different combination of lines within the same diagram, without the need to construct a new diagram.

For a small part of the domain the algebraic procedures are more compact than those for PS diagrams, specifically the calculation of numbers of permutations and combinations for independent events. Table 1 gives the formulas that can straightforwardly be used to calculate the numbers. To obtain the same information with PS diagrams requires diagrams like Figures 28 and 29 to be constructed, using relatively long-winded procedures involving the construction of PS diagrams and the counting of selected events.

However, the PS diagrams do provide an explanation as to why the numbers of permutations and combinations are the size they are in terms of the processes that generate the possible outcomes. The algebraic formulas in Table 1 do not, and to obtain such an explanation under the traditional algebraic approach would require the construction of an outcome table such as Figure 8. These things are consequences of the satisfaction of three semantic transparency criteria for effective representations by the PS diagrams but not by the algebraic representation.

Thus, for problem solvers with the same level of competence, over a broad range of problem types, it is predicted that the solutions that users of PS diagrams produce will be shorter, and contain fewer manipulation errors, than the solutions generated by problem solvers under the conventional approach. In turn, this may enhance learning because procedures with few steps are simple to learn and more of the learners attention can be focussed on the conceptual content of what they are doing, rather than concentrating largely on executing a complex procedure.

Clearly, there are some trade-offs to be made among the characteristics in the design of a representation. As conceptual learning was the particular focus in the development of PS diagrams, rather than pure ease of use of the representation, satisfying the three semantic transparency guidelines had a slightly higher priority than the three plastic generativity guidelines. Nevertheless, PS diagrams appear to meet all six guidelines at least as well as, and in most cases far better than, the representations of traditional algebraic approaches.

DISCUSSION

It is predicted that PS diagrams will be a better representational system for problem solving and learning than the traditional algebraic approach, based on the differences between the systems on each of the six characteristics/guidelines for effective representations. This section considers more generic issues about representations for learning and about the nature of probabilistic reasoning, which build upon the representational analysis in the previous sections. The various claims and observations made in this section are consistent with, and advance the general thesis about, the structural properties of representations significantly determining how effectively they support learning.

Utility of the design criteria for LEDs

The development of PS diagrams provides some evidence of the potential utility of the criteria for the design of LEDs listed above. It is not claimed that the criteria are a necessary and sufficient set for the design of effective LEDs for any domain. Rather, the possibility of specifying such criteria, which placed useful constraints on the successful development of a novel LED with potential to support learning in a complex domain, lends weight to the main thesis of the paper. The features of PS diagrams that were eventually selected to satisfy each of the five criteria are considered in turn.

The need for simple mappings from the things in the domain to the elements in the PS diagrams (criteria Ca), was satisfied by the use of line segments. Each line segment (usually) stands for one event or outcome and its length represents the probability of the event in a linear fashion. The laws of probability were then encoded using simple geometric and spatial relations (Cb). Sums and differences of probabilities are the overall lengths of segments and multiplication of probabilities is achieved by scaling lines with respect to each other. Different cases are apparent from the structure of the diagrams (Cc), because the configuration of the diagram is determined by the relation that is encoded (e.g., Figures 19 and 21).

The fourth criterion (Cd), that specified that the first three criteria should be met at different levels in the domain, is satisfied. For instance, there is one row for each trial in independent situations (Ca). Relations between independent events are captured by simple diagrammatic constraints (Cb) involving the scaling of the lines and the various ways segments may be associated across trials for unlinked and linked events. Alternative situations have quite different structures (Cc), as discussed at length above.

The constraints at the different levels are largely compatible (criterion Ce) with, for instance, the probability of a combination of events equaling the overall length of the lines for the permutations that constitute it. However, there is a minor problem with this as more than one line segment stands for the same event in a particular trial (e.g., Figures 12, 13), which is contrary to criterion Ca at a local level. But this compromise is necessary to allow different permutations and combinations to be distinguished at the more general level in the domain (e.g., Figures 28, 29).

The various trade-offs that had to be made between the criteria indicates that the representational issues are not only concerned with differences between different classes of representations, such as diagrammatic versus sentential. Consideration must also be given to the precise organization of elements in the structure of the notation. This is further illustrated by the reasons for rejection of area as a basis for measures of probability in the development of PS diagrams. The use of bounded regions on the plane to represent events and area for magnitudes of probabilities has been advocated by some (e.g., Armstrong, 1981; Dahlke & Fakler, 1981). This is equivalent to taking areas in Venn diagrams to represent probabilities in addition to sets. For example, in Figure 1 the chances of the conjunction of A□B□C being true is about the same as obtaining A□B□~C, as area 2 is about the same size as area 1.

However, this general approach was rejected for three reasons. First, Venn diagrams are cumbersome to manipulate, as illustrated by the difficulty of extending Figure 1 to show all the possible combinations when another set is added, Figure 2. Second, the mappings from things in the domain to elements of the diagram are often complex (failure to meet criterion Ca). It is sometimes hard to pick out an area for a particular event in a diagram with many overlapping regions. Third, a simple means of encoding the relations among independent events is not feasible (failure of Cb), as demonstrated by the problem of representing independent events using Venn diagrams, discussed above (Figures 4 and 5). A diagram is needed for each trial of independent situations with, for instance, some complex arrangement of arrows to indicate the linking relations between different areas in the diagrams. These failures to meet criteria Ca and Cb occur with respect to both the set theoretic and the probabilistic aspects of the topic, and so implies a failure of criterion Cd that concerns the application of the other criteria of the different levels of a domain.

Ichikawa (1989) presents a 'roulette representation' as a visual model for Bayesian problems. The representation has some of the same basic methods to encode relations as PS diagrams but uses an underlying circular metaphor that is similar to the notion of probability space in PS diagrams. Employing an equivalent approach to representational analysis to that used in the previous paragraphs on the area form of representation, a comparison of PS diagrams and Ichikawa's representation clearly predicts that the use of concentric circles for the trials in a situation is likely to be a substantial disadvantage of the roulette representation.

The benefits of the linear representation in PS diagrams, rather than a circular or an area based system, has been anticipated by the pictorial analog for a specific type of Bayesian problem that was invented by one of the participants in Gigerenzer and Hoffrage’s (1995) studies. The pictorial analog is equivalent to the PS diagram in Figure 30 and the algorithm given by Gigerenzer and Hoffrage for generating the pictorial analog is restricted to the class of problem given in Figure 30. PS diagrams, however, constitute a fully specified representational system covering the full range of probabilistic relations and situations.

The contrast between PS diagrams and other diagrammatic systems for probability highlights the importance of detailed representational analysis in the evaluation of different approaches to the learning of a given domain.

Supplementary representations

The traditional algebraic approach to probability theory was developed over an extended time in the history of mathematics. So it is reasonable to speculate that some aspects of the overall approach will have evolved to cope with the cognitive demands of solving problems and learning the domain, over and above the formal mathematical requirements of the topic. Such developments will include those aspects which address difficulties inherent in the representations themselves. A good example is the need for supplementary representations to support the basic algebraic representation.

Algebra does not provide the means to simply keep track of the interactions of events and possible outcomes, as information needed for problem solving from different levels of abstraction or from different perspectives is typically distributed over many different equations. Substantial cognitive effort is required to remember which terms and which relations are required in, for instance, the solution of the Cab problem when the algebraic representation is used alone. This difficulty can be explained in terms of the analysis of representations provided by Larkin and Simon (1987). Algebra is a <u>sentential</u> representation that employs a method of indexing information among the expressions that is essentially arbitrary. To find particular expressions containing the required information for a particular inference requires significant amounts of search through the list of the previously generated expressions. The order of the expression in the list reflects the sequence in which they were produced rather than anything directly to do with their information content. Further, a cumbersome process of symbolically matching labels in each expression is required to find the target expression. In contrast, Larkin & Simon (1987) argue that diagrams are computationally more efficient than informationally equivalent sentential representations, because they use locational indexing of information. Information that is often used together is likely to be found at adjacent locations in a diagram. This reduces the amount of search in problem solving and alleviates the need for laborious symbolic matching of labels. Thus, it is not surprising that supplementary representations introduced under the traditional algebraic approach with the role of recording and ordering critical information are tree and other diagrams.

The introduction of supplementary representations, and specifically diagrammatic ones, is a good example of how the cognitive difficulties due to the formal structure of a representation requires modification to the overall system of representations to mitigate those difficulties. On a larger scale, the introduction of PS diagrams aims to eliminate many of the inherent problems of the traditional algebraic approach.

Internal and external separability of dimensions

Zhang and Norman’s approach to representational analysis considers how different dimensions of information in a domain are represented using different visual-spatial properties of notational systems (Zhang & Norman, 1994a, 1994b; Zhang, 1997). In particular they highlight the importance of the external separation of the dimensions of information using different properties of the external representation. Lack of external separability means additional information must be retrieved from the internal representation, which increases the difficulty of a task. If different dimensions of the information are externally distributed among the quantity/size, position, shape or colour of elements in the external representation, then external separability will be achieved. The domains they have studied include isomorphs of the Tower of Hanoi and Tic-Tac-Toe, and numeration systems. In the Arabic numeration system shape and position are the visual properties respectively used to represent the base dimensions (i.e., digits '0' to '9') and the power dimension (i.e., right to left 'column' position of digits). By contrast, the

Greek system uses shape to represent both the base and power dimensions, with severe consequences in tasks such as large number multiplication.

This approach to representational analysis can be applied to PS diagrams and the traditional algebra approach. In the algebraic representation, the shape and position of visual properties are used for a mixture of important dimensions of information, as considered in Equations 18-20. Shape is used to represent elements of the domain (i.e., events and probabilities) and it is used for relations between elements (e.g. shape of the 'given' or division symbols, '|' and '/'). Position also has a role in expressing relations (e.g., the order of the alphanumeric symbols in '$S_1|S_2$'). Set theory and probability perspectives are coordinated using both shape and position in the form of parentheses to contain the set expression following the 'P' in the expression for probability terms.

Several properties of diagrammatic elements are used in PS diagrams. The length of a line represents magnitudes of chance, under whichever measure is being considered. Separate line segments indicate different sets and their relative positions indicate such relations as their disjointedness (e.g., overlapping or not) or occurrence on different trials (i.e., vertical separated groups of horizontal lines). Shape and quantity of elements also plays a role. When line segments are labelled, the shape of the labels carries information identifying particular things in the problem situation. In Figures 28 and 29 the quantity of columns for similar sequences of events are counted in the computation of permutations or combinations. The overall patterns of lines is also used to encode different relations, such as the two different set commutativity relations (Figure 10) or the patterns for various probability relations (e.g., Figure 19).

Although it appears, superficially, that PS diagrams may be the better representational system, because it uses the greater variety of visual spatial properties, under a Zhang and Norman’s representational analysis neither system possesses straightforward mappings of particular dimensions of information into unique visual-spatial properties. This lack of external separation of dimensions of information means that there is little basis for considering that one representation would be better than the other. Similarly, this approach would not predict that LEDs for the domains of particle collisions and electricity would be better than traditional approaches using algebra. However, there is evidence that such a differentiation in the effectiveness of LEDs and algebra exists (Cheng, 1996c, 1999b). This is not to claim that Zhang and Norman’s hypothesis about the external separation of dimensions of information is wrong. Rather, it appears that its scope is limited to relatively simple representational systems for domains that have just a few orthogonal dimensions of information on one level, such as the numeration systems and puzzles studied by Zhang and Norman (Zhang & Norman, 1994a, 1994b; Zhang, 1997). PS diagrams, and the other complex representations mentioned, have many dimensions of information arranged in a more hierarchical structure.

The notion of the separability of dimensions may nevertheless be applied to complex domains by identifying whether different representational techniques are used for different dimensions at the same general level within the domain. These techniques need not be simple visual-spatial properties, but may be different classes of geometric or spatial relations. For example, with respect to the different classes of probabilistic situations, Figure 3, PS diagrams have greater external separation of the dimensions, because the representation uses different methods of encoding for each dimension, whereas the algebraic approach does not.

There is a similarity between the notion of the external separability of dimensions of information and the idea of locally heterogeneous representations of concepts, which is one half of the second characteristic in the semantic transparency set proposed above (G2). This emphasizes the point that for complex representational systems, higher level diagrammatic properties of such systems are to be considered when assessing the effectiveness of representations. Simple analysis that match basic visual-spatial properties to simple dimensions of information may be suitable for straightforward cognitive tasks in simple domains, but for complex problem solving and learning in substantial domains characteristics like those proposed here are necessary.

Clarifying the conceptual structure of probability theory

Whilst designing the PS diagrams, reference was made to the algebraic formulations of the laws as the set of relations that needed to be covered by the system. However, some major conceptual difficulties were encountered when thinking about the conceptual relation between the two dimensions: (1) independent versus not independent events, and (2) disjoint versus not disjoint events. Several iterations on the basic design of PS diagrams were required to achieve consistent interpretation of the constraints of the domain with the satisfaction of the design criteria, especially the criterion of compatibility of the constraints over different levels in the domain (Ce). During this work it was realized that there was a lack of conceptual coherence in the way most textbooks discuss these dimensions, which had to be clarified before the design process would produce a consistent and coherent system of Law Encoding Diagrams for the domain.

In textbooks on probability, the concepts relating to the two dimensions are usually introduced with examples and followed by definition in terms of Equations 5, 7, 9 and 10. There is typically no explicit consideration of the conceptual relations and distinctions among the dimensions. Thus, it is possible on first sight to gain the impression that these two dimensions are orthogonal. They are not. In the conceptualization of the domain established for the design of PS diagrams, shown in Figure 3, the second dimension, disjoint versus joint events, is a sub-dimension of the dependent events. The primary distinction in Figure 3 is between independent versus dependent events, and on this basis it was realized that quite different methods were needed to encode these different classes of situations. Thus, the general use of symbols from set theory for relations was abandoned, and new symbols (&, v, «) for independent processes introduced. Further, given these classes of situations, new terms for the relation between independent events were introduced ('linked' and 'unlinked') to avoid confusion with relations within dependent events ('joint' and 'disjoint'). This distinction is also one that is not clearly made in probability textbooks.

A possible explanation for this lamentable state of affairs in many probability textbooks is poor semantic transparency of algebra as a representation for probability. The algebraic conceptualization does not, in itself, make these conceptual distinctions explicit, so instructors who implicitly understand the underlying structure of the domain may not consciously appreciate the need to address the distinctions in their writing. When they do acknowledge this issue, such as in the selection of sets of exercises, the typical use of algebraic expression of the laws does a poor job of making plain the inherent structure of the domain.

Again, the pivotal role of the structure of representations in themselves for complex cognitive processes is highlighted.

Informal reasoning and internal representations

There has been extensive work on the nature of peoples' informal probabilistic reasoning, which has asked whether such reasoning is rational. Using tasks such as Monty's dilemma, the Cab problem and the Problem of the Three Prisoners many studies seem to demonstrate that peoples’ informal reasoning in probability is flawed (e.g., Kahneman, Slovic and Tversky, 1982; Shimojo and Ichikawa, 1989; Falk, 1992; see Shaughnessy (1992) and Evans (1992) for reviews). People do not conform to the canons of probability theory, because of the biases they possess, or because they use informal heuristics or subjective theorems. However, others (e.g., Cosmides and Tooby; 1996) have argued that much of the counter-intuitiveness of the problems is due to the particular framing of the problems rather than any inherent inability of people to deal rationally with probabilities.

Addressing these issues in detail is beyond the scope of this paper, but the representational implications of the work are relevant here. The studies on informal probabilistic reasoning have dealt with the internal mental processing of information given in the problems, by focusing on biases and heuristics. Johnson-Laird (1994) has proposed that his theory of mental models will be extendible to probabilistic thinking. The focus of the present work clearly differs in that it is, in addition, dealing with the role of external representations. There are benefits of using an external representation or doing display-based problem solving (Larkin, 1989; Scaife and Rogers, 1996), such as reductions of working memory demands and the off-loading of some mental computation onto the representation.

A good example of this is seen in the Figure 31, the PS diagram solution to Monty's dilemma, in which the many pieces of information are neatly organized. Further, Shafir (1994) suggests that one reason probability problems are hard is that they often include disjunctions, which people are typically reluctant to think through. Faced with a disjunction people tend to travel down just one branch and suspend judgement when they reach a node. This is often the case with Monty's dilemma. PS diagrams may alleviate this difficulty, because each drawing represents the whole probability space, so missing combinations of events are readily identified. Figure 25 shows complete PS diagrams with all the possibilities enumerated whilst Figure 24 shows partial diagrams. Tree diagrams are likely to be less effective because they require the reasoner to think through the disjunctions in the first place before the tree diagram can be drawn. Unlike PS diagrams, there is no external indication of how completely all the branches have been explored.

Cosmides and Tooby (1996) and Gigerenaer and Hoffrage (1995) present evidence to argue that the apparent difficulties that people have with information in probabilistic reasoning is due, in large measure, to the epistemic ontology of the perspective implicitly adopted in the framing of the problems; i.e., in terms of degrees of belief or chances of single events. When the same problems are rewritten to emphasize frequencies or numbers of events under the aleotoric ontology, peoples' judgements approximated well to the normative laws of probability. This implies that for learners of probability the frequentist/aleotoric perspective should be emphasized and transition to a Bayesian/epistemic conception has to be treated with care. Textbooks on probability typically adopt the epistemic ontology and usually deal tangentially with aleotoric approaches. The notation for relative frequencies is somewhat more cumbersome than the common notation of the epistemic perspective (because two quantities must be given to state a frequency). If those advocating aleotoric oriented approaches are correct, then PS diagrams may have an advantage over the traditional algebraic approach, because PS diagrams can equally well be interpreted in aleotoric as well as epistemic terms, or even both simultaneously.

The use of algebra as an external representation, as opposed to informal verbal reasoning, has implications for internal cognitive processing that extend beyond basic claims about the use of the representation as a form of external memory. A different representation will mean that an alternative set of internal representations and processes will be invoked. Ideally, when the algebraic representation is correctly used, reasoning will conform to the canons of probability theory and problems due to biases and associated with the use of heuristics will be circumvented. However, such problems are manifest in the reasoning of people who have studied probability theory (Shaughnessy, 1992). This may, in part, be attributed to the inherent representational problems of the traditional algebraic approach. The representation does not constitute a semantically transparent system nor does it have plastic generativity. Thus, on meeting impasses, problem solvers often find it easier revert to informal mental inferences rather than struggle with the difficult external representation.

In the same vein, the internal cognitive counterparts to the external aspects of PS diagrams are likely to be quite different to the internal representation and processes used for the traditional algebraic approach. PS diagrams, being LEDs, will in this regard be similar to other LEDs. The framework of schemas for LEDs (Cheng, 1999a), mentioned above, proposes that there are different classes of schemas for concepts at different levels of abstraction and for different levels of complexity of a domain. LEDs for kinematics and dynamics (Cheng, 1999a) and LEDs for electricity (Cheng, 1988a, 1999b) appeared to be acquired and processed as perceptual chunks, thus it is plausible to predict that the internal representations that will underpin learning with PS diagrams will also be some form of perceptual schemas.

CONCLUSIONS

The main thesis of this paper is that in conceptual learning, the nature of the chosen representations significantly determines how easily concepts are learnt and the structure of the network of concepts that is acquired. Analyzing the ontological, structural and functional properties of representations is essential to understanding how they can effectively support learning or can pose substantial formal and cognitive hurdles for the learner to overcome. Here, support for the thesis was provided by: (i) the development of a novel representational system for probability theory using design criteria for LEDs and general characteristics of effective representations; (ii) the analysis of the potential of the new system in comparison to the traditional algebraic approaches to the domain; (iii) the explication of selected representational issues.

This work extends previous work in three ways. First, representational analysis is applied to conceptual learning in addition to problem solving. Second, complex representation systems for complex domains rich in conceptual structure are being considered. Alternative characteristics or properties of representations are needed to explain what makes an effective representation in relation to levels of abstraction, alternative perspectives, and extensive processing of external expressions not found in less substantial domains. Third, the identified characteristics of effective representations have not only been used to analyze representations, but have in effect been operationalized in the design of a new representation. This provides a stringent test of the coherence and utility of the characteristics.

ACKNOWLEDGEMENTS

The work was supported by the U.K. Economic and Social Research Council through the Centre for Research in Development, Instruction and Training.

My gratitude goes to the members of the Centre who have encouraged me in this research, in particular David Wood. Thanks also to David Peebles who made comments on an earlier draft of this paper and to Jo Cheng for the arduous task of proofreading.

REFERENCES

Ainsworth, S. A., Bibby, P., & Wood, D. J. (1997). Information technology and multiple representations: new opportunities - new problems. Journal of Information Technology for Teacher Education, 6(1).

Armstrong, R. D. (1981). An area model for solving probability problems. In A. P. Shulte (Ed.), Teaching Statistics and Probability (pp. 135-142). Reston, VA: National Council of Teachers of Mathematics.

Austin, J. D. (1974). An Experimental Study of the Effects of Three Instructional Methods in Basic Probability and Statistics. Journal of Research in Mathematics Education, 5(3), 146-154.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. Journal of Experimental Psychology: Applied, 4(2), 75-100.

Cheng, P. C.-H. (1994). An empirical investigation of law encoding diagrams for instruction. In Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. (pp. 171-176). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cheng, P. C.-H. (1996a). Law encoding diagrams for instructional systems. Journal of Artificial Intelligence in Education, 7(1), 33-74.

Cheng, P. C.-H. (1996b). Scientific discovery with law encoding diagrams. Creativity Research Journal, 9(2&3), 145-162.

Cheng, P. C.-H. (1996c). Learning Qualitative Relations in Physics with Law Encoding Diagrams. In G. W. Cottrell (Eds.), Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society (pp. 512-517). Hillsdale, NJ: Lawrence Erlbaum.

Cheng, P. C.-H. (1998a). A Framework for Scientific Reasoning with Law Encoding Diagrams: Analysing Protocols to Assess Its Utility. In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 232-235). Hillsdale, NJ: Lawrence Erlbaum.

Cheng, P. C.-H. (1998b). Some reasons why learning science is hard: Can computer based law encoding diagrams make it easier? In B. P. Goettl, H. M. Halff, C. Redfield, & V. Shute (Eds.), Intelligent Tutoring Systems (pp. 96-105). Berlin: Springer-Verlag.

Cheng, P. C.-H. (1999a). Networks of Law Encoding Diagrams for Understanding Science. *European Journal of Psychology of Education*, **14**(2), 167-184.

Cheng, P. C.-H. (1999b). *Electrifying representations for learning: An evaluation of AVOW diagrams for electricity*. (Technical No. 62). ESRC Centre for Research in Development, Instruction & Training.

Cheng, P. C.-H., Cupit, J., & Shadbolt, N. R. (in press). The foundations of knowledge acquisition from graphs. International Journal of Human Computer Studies.

Cheng, P. C.-H., & Simon, H. A. (1992). The right representation for discovery: Finding the conservation of momentum. In D. Sleeman & P. Edwards (Eds.), Machine Learning: Proceedings of the Ninth International Conference (ML92) (pp. 62-71). San Mateo, CA: Morgan Kaufmann.

Cheng, P. C.-H., & Simon, H. A. (1995). Scientific Discovery and Creative Reasoning with Diagrams. In S. Smith, T. Ward, & R. Finke (Eds.), The Creative Cognition Approach (pp. 205-228). Cambridge, MA: MIT Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121-152.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Ed.). (1988). The Nature of Expertise. Hillsdale, N.J.: Lawrence Erlbaum.

Cleveland, W. S., & McGill, R. (1985). Graphical perception and graphical methods for analysing scientific data. Science, 229, 828-833.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. Cognition, 58, 1-73.

Dahlke, R., & Kakler, R. (1981). Geometric probability. In A. P. Shulte (Ed.), Teaching Statistics and Probability (pp. 143-153). Reston, VA: National Council of Teachers of Mathematics.

Ericsson, K. A., & Smith, J. (1991). Toward a General Theory of Expertise. Cambridge: Cambridge University Press.

Evans, J. S. B. T. (1992). Biases in thinking and judgement. In M. T. Keane & K. Gilhooly (Eds.), Advances in the Psychology of Thinking (pp. 95-125). Hemel Hempstead, Hertfordshire: Harvester-Wheatsheaf.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. Cognition, 43, 197-223.

Fischbein, E., & Schnarch, D. (1997). The Evolution With Age of Probabilistic, Intuitively Based Misconceptions. Journal of Research in Mathematics Education, 28(1), 96-105.

Garfield, J., & Ahlgren, A. (1988). Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research. Journal of Research in Mathematics Education, 19(1)44.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, **102**(4), 684-704.
Giarrantano, J. C., & Riley, G. (1989). Expert Systems: Principles and Programming. Boston, MA: PWS-Kent.
Glasgow, J., Narayanan, N. H., & Chandrasekaran, B. (Ed.). (1995). Diagrammatic Reasoning: Cognitive and Computational Perspectives. Menlo Park, CA: AAAI Press.
Goodman, N. (1968). Languages of Art: An Approach to a Theory of Symbols. Indianapolis, IN: Bobbs-Merrill.
Hacking, I. (1975). The Emergence of Probability. Cambridge: Cambridge University Press.
Ichikawa, S. (1989). The role of isomorphic schematic representation in the comprehension of counterintuitive Bayesian problems. Journal of Mathematical Behaviour, 8, 269-281.
Johnson-Laird, P. N. P. (1994). Mental models and probabilistic thinking. Cognition, 50, 189-209.
Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgement Under Uncertainty. Cambridge: Cambridge University Press.
Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. Cognitive Science, 14, 511-550.
Kosslyn, S. M. (1989). Understanding charts and graphs. Applied cognitive psychology, 3, 185-226.
Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Cognitive Psychology, 17, 248-294.
Krauss, L. M. (1994). Fear of physics: a guide for the perplexed. London: Cape.
Kreyszig, E. (1983). Advance Engineering Mathematics (5th ed.). New York, NY: John Wiley.
Larkin, J. H. (1989). Display-based Problem Solving. In D. Klahr & K. Kotovsky (Eds.), Complex Information Processing: The Impact of Herbert A. Simon (pp. 319-341). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.
McColl, J. H. (1995). Probability. London: Edward Arnold.
Newell, A., & Simon, H. A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.
Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), Cognition and Catergorization. (pp. 259-303). Hillsdale, N.J.: Lawrence Erlbaum.
Peterson, D. (Ed.). (1996). Forms of Representation. Tunbridge Wells: Intellect.
Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Eds.), Artificial Intelligence and the Future of Testing (pp. 73-126). Hillsdale, NJ: Lawrence Erlbaum.
Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? International Journal of Human-Computer Studies, 45, 185-213.
Schwarz, B., & Dreyfus, T. (1993). Measuring integration of information in multirepresentational software. Interactive Learning Environments, 3(3), 177-198.
Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. Cognition, 50, 403-430.
Shaughnessy, M. J. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Eds.), Handbook of Research on Mathematics Teaching and Learning (pp. 465-494). New York, NY: MacMillan.
Shimojo, S., & Ichikawa, S. (1989). Intuitive reasoning about probability: theoretical and experimental analyses of the "problem of three prisoners". Cognition, 32, 1-12.

Skemp, R. R. (1986). The Psychology of Learning Mathematics (2nd ed.). Harmondworth, Middlesex: Pelican.

Stenning, K., & Oberlander, J. (1995). A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cognitive Science, 19(1), 97-140.

Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. Cognitive Science, 21(3), 305-350.

White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. Cognition and Instruction, 10(1), 1-100.

Zhang, J. (1997). The nature of external representations in problem solving. Cognitive Science, 21(2), 179-217.

Zhang, J., & Norman, D. A. (1994a). A representational analysis of numeration systems. Cognition, 57, 271-295.

Zhang, J., & Norman, D. A. (1994b). Representations in distributed cognition tasks. Cognitive Science, 18(1), 87-122.