Diagrammatic Re-codification of Probability Theory:
A Representational Epistemological Study

Peter C-H. Cheng (p.c.h.cheng@sussex.ac.uk)
Department of Informatics/COGS
University of Sussex, Falmer, Brighton, BN1 9QH, UK

Abstract

It is claimed that the current representations used for probability theory provides a poor codification of that knowledge. The limitations of the representations and how they encode the knowledge causes conceptual difficulties and makes problem solving difficult. Probability Space diagrams constitute a new representational system that provides a simpler and more coherent codification of probability theory, which effectively supports problem solving. The two approaches are contrasted, which demonstrates that the effects of representations on advanced forms of cognition extend beyond problem solving and impacts on the conceptual understanding and learning of bodies of knowledge.

Introduction

Inventing new representational systems is one of the pinnacles of human intellectual endeavors. The history of science and mathematics is replete with examples of the critical role of the invention of new representations in discovery (e.g., Cheng & Simon, 1995). Effective representational systems are fundamental to the progression and accumulation of knowledge and vital to its acceptance and use by future generations. For instance, the Hindu-Arabic numeration system is a prime example of an effective representation (Zhang & Norman, 1994). It competed with dozens of other systems (Roman, Greek, Mayan, Chinese, etc) that were invented over thousands of years, but through a process akin to natural selection, it is now dominant worldwide.

In contrast, the representational systems used in most technical fields have only existed for a relatively short time and have had few, if any, competitors. The way many subjects are currently codified appears to be due to historical accident rather than deliberated design – they do not have the evolutionary pedigree of the Hindu-Arabic numeration system. Thus, the representational systems that codify knowledge in some subjects may be far from optimal, not only in terms of their suitability for problem solving but in their support for learning.

The importance of the representations for higher cognition is underscored by established findings in cognitive science. Work on problem isomorphs has shown that the difficulty of solving a problem can vary by over an order of magnitude with different representations of the same task (Kotovsky, Hayes, & Simon,1985). Alternative informationally equivalent representations may demand quite different amounts of computation (Larkin & Simon, 1987). Zhang (1997) argued that there is a form of representational determinism in which the nature of the representation used substantially determines what information can be perceived, what processes can be activated, and what structures can be discovered.

It has been argued that such representational effects impact the process of scientific discovery (Cheng, 1996). Previous work has also shown that by re-codifying knowledge of complex conceptual domains, using alternative representational systems, it is possible to significantly enhance learning. For the domain of particle collisions in physics, using the diagrams that the original physicists invented in order to discover the laws of momentum and energy conservation, students gained a better understanding compared to others using the modern algebraic approach (Cheng, 1999a). By inventing a new diagrammatic representation that encodes the laws of basic electricity (Cheng, 2002) it was found that students could obtain a better conceptual understanding of that apparently difficult topic compared to students learning using the current algebraic approach (Cheng, 2002; Cheng & Shipstone, 2003).

The term *representational epistemology* may be used to denote studies that examine the fundamental role of representational systems in the advanced cognition (problem solving, learning, discovery) of complex knowledge-rich domains. There are at least four stages in such studies. First, the content and problem classes of a target domain are analyzed to reveal the underlying conceptual structure of the knowledge, which includes the ontologies, perspectives, scale levels, laws, models, prototypes and extreme cases of the domain. Second, the current domain representations are examined to understand how effectively, or otherwise, they support conceptual understanding and problem solving, for instance using task analyses. Third, a new representational system is invented (or a historical representation revived) that provides an alternative encoding of the conceptual structure of the domain and an alternative approach to problem solving. Fourth, empirical evaluations to compare problem solving and learning with the new representations and the conventional representations are conducted.

One aim of such representational epistemological studies is to discover principles of representational design (e.g., Cheng, 1999a, 2002). Such principles will support the further analysis of the effectiveness of extant representations and may be used as heuristics for the design of better representational systems for other domains.

This paper reports a representational epistemological study on basic probability theory. It is claimed that the existing representations used to encode the knowledge of probability is poor. Much of the difficulty of understanding, problem solving and learning this topic is caused by inherent limitations of the current representations. Support for

this claim comes from an analysis of the difficulties of the conventional approach (next section) and by contrasting it with a new representation – *Probability Space (PS) diagrams* – that has fewer limitations (following section).

The study addressed probability theory at upper high school level. To give a sense of this level, consider a problem that will be used as running example throughout this paper – the Card problem.

> There are three cards in a bag. One card has both sides green, one card has both sides blue, and the third card has a green side and a blue side. You pull a card out, and see that one side is blue. What is the probability that the other side is also blue?

Students often find this problem counter intuitive and fail to solve it correctly (Shaughnessy, 1992). It reflects the difficulties of the conventional approach well. One indication of the value of PS diagrams is that they provide simple explanations of such apparently paradoxical problems. (The answer is 2/3.)

Conventional Approach

The first stage of this representational epistemological study was to characterize the conceptual structure of the domain, identify the extant representations used and the typical procedures for problem solving. For probability theory this was done by examining established, well regarded, student texts (e.g., Greer, 1995; McColl, 1992). These cognitive artifacts presumably reflect how expert teachers consider the topic should be understood.

The identification of the representations was straightforward. The texts all included Venn diagrams, tree diagrams, tables of outcomes, algebra and, of course natural language descriptions. In the more advanced texts set theory notation, contingency tables and probability density distribution graphs were found.

The conceptual structure of a domain concerns the major conceptual dimensions of that domain – the primary meaningful distinctions, categories of things, universal invariants and overarching relations. This is not equivalent to a statement of the axioms of probability nor is it the same as the ontological underpinnings argued over by philosophers (e.g., Hacking, 1975). Rather the conceptual structure comprises the foundational components of knowledge upon which the domain is based that student must master to fully understand the domain.

Characterizing the conceptual organization of probability theory as portrayed by the students texts was a surprisingly difficult task. Although the division of topics into chapters and sections in the texts suggested possible structures, these were not consistent across texts and left many aspects largely implicit. Nevertheless, Table 1 shows one possible interpretation, with the main conceptual dimensions and forms of relation shown using the nested columns and rows. The numbers of events and whether they are joint, dependent and equi-probable, or otherwise, are fundamental to this encoding of the knowledge of probability theory. Although equations are presented in the cells to stand for the specific instances of concepts or propositions, examples of the other representations listed above could have been used to illustrate the concepts.

Similarly, extracting the approach to problem solving embodied in the texts was also difficult, largely because the differences between solution procedures associated with

Table 1. Conceptual structure of probability theory: laws for the main conceptual distinctions and relation types.

Singe events (A - event, U - universal set)			
(T1[*]) $P(U)=1$	(T2[*]) $0 \leq P(A) \leq 1$	(T3) $P(A)=1-P(\sim A)$	(T4) $P(A)=P(A|U)$ [@]

Multiple events (A, B - events)			
	Joint		
Relations	Disjoint	Independent	Dependent
Conjunction	(T5) $P(A \text{ and } B)=0$	(T6) $P(A \text{ and } B) =P(A)P(B)$	(T7) $P(A \text{ and } B)=P(A)+P(B)-P(A \text{ or } B)$ (T8) $P(A \text{ and } B)=P(A|B)P(B)$
Disjunction	(T9[*]) $P(A \text{ or } B)=P(A)+P(B)$	(T10) $P(A \text{ or } B)=P(A)+P(B)-P(A \text{ and } B)$ (T11) $P(A \text{ or } B) =1-P(\sim A \text{ and } \sim B)$	(T12) $P(A \text{ or } B)=P(A)+P(B)-P(A \text{ and } B)$
Conditional	(T13) $P(A|B)=0$, $P(B|A)=0$	(T14) $P(A|B)=P(A)$, $P(B|A)=P(B)$	(T15) $P(A|B)=P(A \text{ and } B)/P(B)$, $P(B|A)=P(B \text{ and } A)/P(A)$
Complex	(T16) $P(A|B) =P(B|A)P(A)/P(B)$		

Equi-probable multiple events (N, M - possible outcomes)			
	(T17) $P(i)=1/N$, $\Sigma P(i)=1$	(T18) $P(i,j)=1/(N \cdot M)$, $\Sigma\Sigma P(i,j)=1$	
Permutations[†]		(T19) No. of outcomes $= n^k$	(T20) No. of outcomes $= \dfrac{k!}{(n-k)!}$
Combination[†]		(T21) No. of outcomes $= \dfrac{(n+k-1)!}{k!(n-1)!}$	(T22) No. of outcomes $= \dfrac{n!}{k!(n-k)!}$

Note: * — Axioms of probability theory, @ — Not normally state explicitly, † — k selected objects from an initial set of n objects.

Analyse

A.1) If outcomes disjoint, then consider number of events:

  A.11) if a single event, then consider nature of probabilities of outcomes:

  A.111) if equal probabilities problem, then list items of interest:

  A.1111) if simple problem, then list all items & select target outcome.

  A.1112) if complex problem, then list relevant items & select the target outcome.

  A.112) if unequal probabilities problem, then consider individual probability values:

  A.1121) if simple problem, then list selected items with associated values and select target outcome.

  A.1122) if complex problem, then draw up a 1-way table outcome table and select target outcome(s).

  A.12) if multiple events, then consider nature of the probabilities of outcomes:

  A.121) if equal probabilities, then consider number of events:

  A.1211) if two events, then generate a systematic list or use a 2-way outcome table and select target outcomes.

  A.1212) if many events, then use a recursive tree diagram to enumerate relevant outcomes and select target outcomes.

  A.122) if unequal probabilities, then list relevant outcomes with their probability values and select the target outcomes.

A.2) If events joint, address the relation between events:

  A.21) if events dependent, then consider number of events:

  A.211) if one event, then consider the nature of probabilities of outcomes:

  A.2111) if equal probabilities, then consider number of types of outcomes:

  A.21111) if one type of outcome, then use a Venn diagram.

  A.21112) if two types of outcomes, then use a two-way table.

  A.2112) if unequal prob's, then use a Venn diagram to elaborate set relations and write target outcomes with probability values.

A.212) if sequences of events, then use a tree diagram with dependent branches and values and select target outcomes.

A.22) if outcomes independent, then tree diagram with repeated branches & values, or contingency table, to select the target outcomes.

Calculation

C.1) If events are disjoint, then:

  C.11) if equal probabilities, then count number of outcomes of interest and use T17 to find probability of each outcome.

  C.12) if unequal probabilities problem, then use T5, T9, T13 depending on target relation.

C.2) If events are joint, then consider the number of events and outcomes:

  C.21) if the number of events and outcomes are small, then consider nature of the relations between events:

  C.211) if independent events, then consider nature of probabilities:

  C.2111) if outcomes are equi-probable, then compute number of outcomes of interest and use T18 to find probability of each outcome.

  C.2112) if unequal probabilities, then use T6, T10, T11 or T14 depending on target relation.

  C.212) if dependent outcomes, then consider complexity of dependencies:

  C.2121) if simple dependencies, then use T7, T8, T12 or T15 depending on target relations.

  C.2122) if complex interconnected dependencies, then use Bayes theorem, T16.

  C.22) if events and/or outcomes are numerous, then consider nature of probabilities:

  C.221) if equal probabilities, then use T19-T22 depending on the natures of dependencies and the target relations.

  C.222) if unequal probabilities, then use Bayes' theorem, T16.

Notes: 'T' numbers refer to Equations in Table 1. Rules at the same decimal level are disjunctive alternative (e.g., A.11 or A.12), and rules with increasing decimal places are conjunctive sequences (e.g., A.11 then A.111 then A.1111).

Fig. 1. General problem solving procedure for the conventional approach.

alternative concepts were not made explicitly in the texts. However, focusing on typical quantitative problems, which involve finding the probability of a given situation, it was possible to induce a generic solution approach, as shown in Fig. 1. The approach is broken down into an analysis phase and a calculation phase, each comprising of nested rules. Analysis involves interpreting aspects of a problem in terms of the main conceptual dimensions and identifying the event(s) of interest, usually with the aid of the representations tailored to particular relations or concepts. Calculation involves selection of the equations to compute the probability of the target event. Different organizations of rules to that in Fig. 1 are possible, with the dimensions occurring higher or lower in the decision hierarchy. However, they would not be as consistent with the texts and would likely be as complex.

The difficulties of using Table 1 and Fig. 1 to solve probability problems is well illustrated by an ideal solution to the Card problem, as shown in Fig. 2. It can be broken down into seven steps:

1) Overall, this is a situation in which the events are joint and occur in sequence, thus a tree diagram with dependent branches is drawn, Fig. 1 (A.212).

2) One of three cards can be drawn from the bag at random, so three disjoint events are listed and drawn as branches (A.1111), and labeled with equal probabilities using T17 (C.11)

3) For each of the three cards the color of the visible side depends on the card being chosen, so the one or two possible conditional dependent events are shown by extending the branch or splitting it equally (A.212), and labeling them with their respective probabilities, using T15 (C.2121).

4) The color on the opposite side of each card is a conditional event that depends on which side of that card is already showing, so the branches of the tree are further extended (A.212) and label with unit probabilities, using T15 (C.2121).

5) Each leaf of the tree is the independent result of series of joint events (A.22). The probability of each leaf is calculated separately using T6 (C.2112).

6) Overall initial chance of seeing a blue card is comprised by the disjunction of the disjoint events of the blue sides showing (A.122) using T9 (C.12).

7) Finally, the chance of turning over a blue side is dependent and conditional upon the card and side already showing (A.212). Calculate the probability using T15 (C2121).

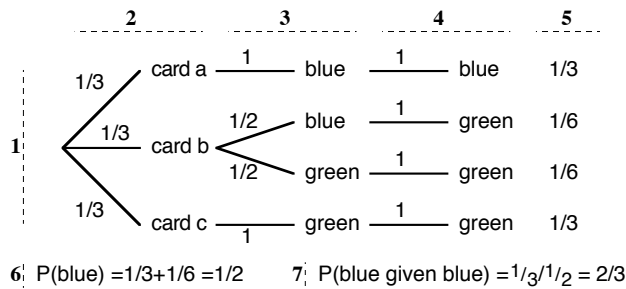The second stage of the representational epistemological study was to consider the effectiveness of the conventional

Fig. 2. Tree diagram for solution to the card problem. Bold numbers with dashed lines are solution steps.

Fig. 3. A schematic PS diagram. E1, E2 – trials/experiments; A, B, C and Q, R, S, T - outcomes of E1 and E2.

approach. From analysis of this and other quantitative problems and examining the structure of Table 1 and Fig. 1, reasons for the difficulties of the conventional representation were identified. There are at least 7 reasons why the current approach is poor. (a) For theoretically distinct concepts, it uses terms that have similar common meanings (e.g., *joint* versus *dependent*). (b) Some terms are ambiguous. For instance, *event* has different meanings, including *experiment/trial* and *outcome*. Examples of the simultaneous use of these different meanings within a single explanation has even been found in some texts. (c) The conceptual structure is not simply a set of orthogonal dimensions but involves nested concepts within an asymmetric organization; for instance, *(in)dependence* occurs under *joint* but not under *disjoint* (Table 1). (d) Many different representations are needed for modeling different classes or aspects of situations. (e) The general procedure for solving simple quantitative problems is complex, with multiple nested disjunctive sequences of rules (Fig. 1). (f) The procedure has to be applied recursively to different parts of the problem, which may involve switches of conceptualization within a single problem (e.g., Card problem). (g) It is often necessary to fully appreciate the inherent form of a problem before beginning so that a solution approach can be chosen. For instance, the Card problem consists of joint independent sequences of dependent events(!). However, the initial random selection of cards may mislead one into seeing this as a disjoint events problem and incorrectly drawing a Venn diagram.

Given these difficulties it is no wonder that probability theory is well acknowledged to be conceptually challenging topic, which is both hard to understand and to learn.

Probability Space diagrams

The third stage of the representational epistemological study was to invent a new representational system that attempts to overcome the limitations of the conventional approach. The representational principles outlined by Cheng (1999a, 2002) were used as heuristics to constrain the search of possible representational schemes. Probability Space diagrams were the result (Cheng, 1999a, Cheng & Pitt, 2003). A key idea in the design of PS diagrams was to provide a representational scheme wit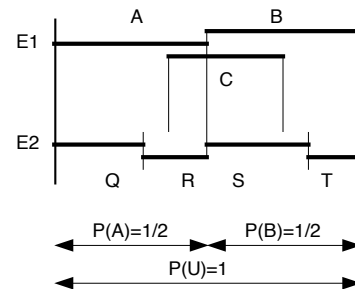h an orthogonal, or distinct, encoding the primary conceptual dimensions that gives a coherent interpretive scheme for the other concepts and relations. It was found that such a scheme could be achieved by taking *trials/experiments* and *outcomes* (of trials) as orthogonal foundational dimensions which were used to defined a *probability space*. Fig. 3 shows a schematic PS diagram. The space encompasses the universe of interest (U) and its width has unit probability by definition. The outcomes of each trial are represented by grouped line segments. The lengths the lines are in proportion to the probability of the represented outcomes. The spatial and geometric relations among the segments encode the axioms and laws of probability theory. For instance in Fig. 3, outcomes A and B are the only mutually exclusive outcomes of trial E, so P(A)+P(B)=1. Overlapping segments represent non-mutually exclusive outcomes (e.g., A and C) and the extent of the overlap is the probability of both occurring.

Relations, or *links*, from one trial to another are encoded by the horizontal alignment of outcomes across the trials. For example, Q and R are possible outcomes in trial E2, if A was the outcome of trial E1. P(A and Q) is given by the length of Q (with respect to U). Thus, conjunctive and conditional relations within and between trials are simply and consistently encoded.

For small numbers of outcomes and trials, PS diagrams encode all of the laws given in Table 1. See Cheng and Pitt (2003) for discussion of the limitations of PS diagrams.

Fig. 4 shows a generic approach to solving quantitative probability problems with PS diagrams, which was derived by using PS diagrams to solve problems and to design curriculums for teaching using the representation. The approach splits problem solving into distinct modeling, interpretation and calculation phases. The modeling phase involves drawing a PS diagram for the problem situation, including one or more sets of lines for the trials and aligning the outcomes across trials to encode the given links or contingencies. The interpretation phase involves identifying the target outcomes and the relations within the diagrammatic model. The final calculation stage involves simple geometry to find the ratio of the length of the target to the appropriate space to give the required probability.

The solution to the Card problem illustrates the procedure, Fig. 5:

Modelling situation:
M.1) If single trial:
  M.11) identify all outcomes and construct line for the trial:
    M.111) if equal probabilities, then segments equal for outcomes.
    M.112) if unequal probabilities, then make segment proportional to their probabilities.
M.2) If multiple trials:
  M.2a) choose an order of trials and consider each trial in turn:
  M.2b) draw first trial using rule M.1
  M.2c) for second and subsequent trials:
    M.2c1) if trials unlinked, then repeat the trial drawn to scale under each outcome of previous trial.
    M.2c2) if trials linked, then draw the next trial under each outcome consistent with local problem constraints:
      M.2c2a) if outcomes are linked across trials, then change the possible set of outcomes.
      M.2c2b) if probabilities linked, then change relative length of outcomes.

Interpretation:
I.a) Identify target outcome(s) or sequences of outcomes
  I.a1) if single trial, then select target outcome.

I.a2) if multiple trials, then consider nature of outcomes of interest:
  I.a21) if particular outcomes in some of the trials are of interest, then select those outcomes.
  I.a22) if outcomes across trials are of interest, then consider sequences down the diagram:
    I.a221) if permutations, then consider segments down diagram in order.
    I.a222) if combinations or conjunctions, then consider columns containing the target in any order.
I.b) select outcome relation of interest:
  I.b1) if outcome/sequence or its complement, then it is the target.
  I.b2) if union, then the target encompasses outcomes/sequences of interest.
  I.b3) if intersection, then the target is the overlap of outcomes/sequences.
  I.b4) if conditional, then identify subspace for target segment.
  I.b5) if permutation, then the target is the desired sequence.

Calculation:
C.1) Find probability by comparing the length of the target to width of the (sub)space.

Notes: Rules ending with a number at the same level are disjunctive alternative (e.g., M.2c1 or M.2c2); rules with successive letters are conjunctive sequences at the same level (M.2a then M.2b); rules with increasing decimal places are conjunctive sequences increasing in detail (e.g., M.2c then M.2c2 then M.2c2a).

Fig. 4. Procedure for problem solving with PS diagrams.

1a) A trial line showing all of the possible color outcomes of selecting cards with either of their sides showing is drawn (rules M.2a and M2b, Fig. 5).

1b) A second trial line is drawn showing how as each card is turned over the color depends on the previously showing color (M.2c2a).

2a) The outcomes for blue sides initially showing in the first trial are highlighted, dashed target line, defining a conditionalising subspace (I.b4).

2b) The outcomes for blue sides showing after the turn of the card are highlighted within that subspace by the second dashed target line (I.a21).

3) The desired probability is the ratio of the lengths of the second to the first target lines (C.1).

This PS diagram is one that a user with some experience with PS diagrams is likely to produce, grouping initial blue sides outcomes together. Such users might omit the outcomes to the right half of the second trial line, recognizing that they are not relevant to the problem. Novice users may not arrange the outcomes so conveniently but this does not
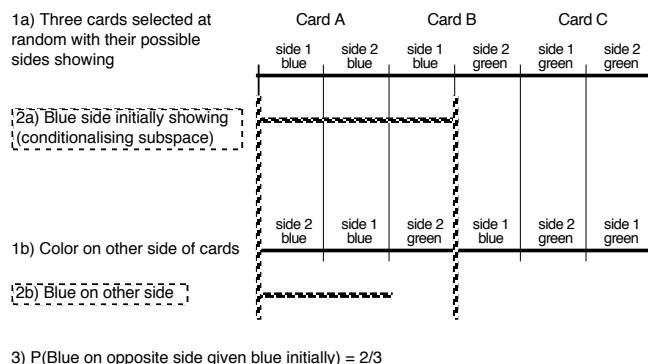
especially hinder the solution as the target outcomes will be dispersed but the linking between each trial and the numbers of outcomes will still be clear.

In contrast to the conventional approach, in both conceptual structure and task analytic terms, PS diagrams constitute a better representational system for 8 reasons. (a) Distinct concepts have been given terms that are compatible with their more common meanings (e.g., *trial*, *outcomes*, *linking*, *overlapping*). (b) Ambiguous use of terms has been avoided (cf. *event*). (c) The conceptual structure is relatively simple with orthogonal primary dimensions (*trials* versus *outcomes*) providing an overarching interpretive scheme within which other concepts can be consistently interpreted (horizontal relations within trials, vertical forms of linking between trials). (d) The PS diagrams approach uses a single representational system. This reduces the cognitive load and learning demands of multiple representations and mapping between them. PS diagrams in effect, combine the functions of Venn diagrams, tree diagrams, set theory notation, contingency and outcome tables, and algebra. (e) The general procedure for solving simple quantitative problems has fewer steps and decision points (compare Figs. 2 and 4). (f) Problems can often be solved in a single pass through the procedure, without the need for recursive application. Thus a clear separation between modeling, interpretation and calculation phases can be maintained in problem solving and instruction. (g) Problem solving can begin without the user fully understanding overall structure of the problem as the modeling phase allows the user to incrementally build a PS diagram. (h) PS diagrams encode a greater proportion of the problem solving information and underlying laws of probability theory using visual properties and geometric and spatial relations, rather than sentential/propositional forms. Thus, more of the benefits of diagrams can be exploited (Larkin & Simon, 1987), including

Fig. 5. PS diagram solution to the card problem

the use of spatial indexing of information to assist search and perceptual operators to aid recognition of patterns, relations and even errors.

Discussion

From the comparison between the two approaches, based on the examination of the structure of the alternative encodings of the knowledge in each approach and the task analyses on quantitative problems, overall it is clear that the PS diagrams provides a better encoding of the knowledge of basic probability theory at the level targeted. Fuller analyses can be found in Cheng (1999b). Further, an experiment in which students learnt probability theory using the two approach, given the same curriculum content and instruction time, has showed that PS diagrams significantly enhanced conceptual learning and problem solving compared to the conventional approach. That empirical study will be reported elsewhere.

Why is the PS diagram approach better? An explanation can be given in terms of the principles used to guide its design. The six principles are concerned either with making the conceptual structure of knowledge readily apparent in the structure of a representational system, *semantic transparency*, or enhancing the efficiency of problem solving, *syntactic plasticity*. The principles are outlined and discussed elsewhere (Cheng, 1999a, 2002).

The focus here is on more general representational epistemological claims. First, the conventional representations used to embody the knowledge of probability theory, as advocated by current student texts, constitute poor codification of that knowledge. The approach provides a complex conceptual structure and an elaborate procedure for problem solving, which needs the support of diverse representations. Second, it was possible to design a new representation that appears to overcome much of the difficulties of the conventional representation. PS diagrams provide a simpler encoding of the knowledge with a more straightforward problem solving approach using a single representational system, in the main. It thus appears that some, perhaps much, of the difficulty of understanding, using and learning probability theory may be caused by the nature of the representations used to encode that knowledge. This provides an alternative perspective on the difficulties of understanding probability theory, which complements accounts that either focus upon the hard concepts at the core of the subject (Shaughnessy, 1995) or the adoption of (in)appropriate ontologies (cf., Gigerenzer & Hoffrage, 1995).

It is well established in cognitive science that alternative representations used in problem solving will dramatically influence the ease, and even the nature of, solution processes (Zhang, 1997). The representational epistemological study presented here provides further support for the claim that representational effects (determinism) can be found not only at the level of problem representation but more generally in how the knowledge of a domain is codified.

References

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

Cheng, P. C.-H. (1999a). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers in Education*, *33*(2-3), 109-130.

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

Cheng, P. C.-H. (2002). AVOW diagrams enhance conceptual understanding of electricity: Principles of effective representational systems. *Cognitive Science, 26*(6), 685-736.

Cheng, P. C.-H., & Pitt, N. G. (2003). Diagrams for difficult problems in probability. *Mathematical Gazette*, **87**(508), 86-97.

Cheng, P. C.-H., & Shipstone, D. M. (2003). Supporting learning and promoting conceptual change with box and AVOW diagrams. Part 1: Representational design and instructional approaches. *International Journal of Science Education, 25*(2), 193-204.

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*, Cambridge, MA: MIT Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, *102*(4), 684-704.

Greer, A. (1992). *A Complete GCSE Mathematics Higher Course* (Third ed.). Cheltenham, UK: Stanley Thornes.

Hacking, I. (1975). *The Emergence of Probability*. Cambridge: Cambridge University Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? *Cognitive Psychology, 17*, 248-294.

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.

Shaughnessy, M. J. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning*, New York, NY: MacMillan.

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

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