71

Designing software for cognitive change: StatPlay and understanding statistics

Geoff Cumming
Mark Zangari

La Trobe University
Bundoora
Australia

Neil Thomason

University of Melbourne
Parkville
Australia

ABSTRACT

We discuss the design of Computer Based Learning Environments and argue that multiple representations are valuable and that learners need guidance and structured activities, even when an exploratory, learner centred approach is adopted. We describe StatPlay, a collection of demonstrations and interactive simulations intended to promote cognitive change and good understanding of central aspects of statistics and probability offering multiple representations. In addition to free exploration challenging tasks—some in game formats—offer structure and guidance to the learner's activities. StatPlay is being developed in Visual C++ for Windows. We describe use of StatPlay by students and a quasi-experiment tracking the development of their sampling concepts.

Main conference themes: learner centred learning
Educational areas: higher education
Study topics: mathematics, humanities/social sciences
Secondary keywords: cognition, learning systems, simulation

INTRODUCTION

We discuss the design of learning environments which can promote cognitive change and rich understanding of difficult concepts.   To illustrate the discussion we describe StatPlay, a collection of demonstrations and interactive simulations intended to help learners understand some statistical concepts which are notoriously difficult to learn, and are widely misunderstood even by many researchers.

Initial assumptions
We adopt the following initial assumptions:

- Learners should have scope for taking initiative and exploring.
- Learner engagement is central; this will be maximized by excellent interface design and challenging activities, including games.
- It is useful to regard learning as conceptual change, in which naive concepts are replaced by 'correct' concepts, meaning those of an expert or an accepted theory.
- Good understanding has been achieved when a learner can explain concepts in several ways, usually including abstract and concrete forms, and can use the concept in practice.

STATISTICAL MISCONCEPTIONS

Misconceptions of statistics and probability are widespread, persistent, resistant to conventional education and have severely damaging consequences [1].   We briefly describe two of the most troublesome faulty intuitions.

Misconceptions of sampling
Tversky and Kahneman [2] presented evidence of a fundamental misconception about randomness: people generally underestimate the amount of variability from sample to sample and overestimate the similarity of a sample to the population it came from.   Equivalently they overestimate the chance that repetition of an experiment will give a similar outcome. Further they fail to realize the crucial role of N, the sample size, in all these judgements. Here is Tversky and Kahneman's classic example:

> Suppose you have run an experiment on 20 subjects, and have obtained a significant result— (z = 2.23, p < .05, two-tailed). You now ... run an additional group of 10 subjects. What do you think the probability is that

the results will be significant, by a one-tailed test, separately for this group? [2, p. 105]

Tversky and Kahneman found the median answer given by research psychologists was 0.85, whereas 0.48 is a much more defensible estimate: the subjects greatly overestimated the agreement likely between successive experiments with small sample sizes.

Such erroneous judgements lead to over-interpretation of discrepancies between experiments which in fact reflect only sampling fluctuations. In this way wrong conclusions are often drawn when large bodies of research are reviewed; Schmidt [3] argued that whole fields of psychological enquiry, and practical application, have been severely damaged in this way.

Faulty intuitions about sampling lead to severe under-estimation of the width of confidence intervals: researchers believe their experimental results are much more accurate than is justified. Such intuitions also underlie misconceptions of statistical power, with the result that much research effort is wasted conducting low-powered experiments which have only a poor chance of detecting any effect that may exist.

Misconceptions of statistical significance

Misunderstanding of statistical significance and hypothesis testing has been discussed by Gigerenzer [4], Oakes [5] and many others. One problem is confusion between the statistical significance of an effect and its size or importance. Another is interpretation of a lack of statistical significance as meaning the null hypothesis is true. Yet another is the belief, discussed above, that repeating an experiment with a statistically significant result is very likely again to give statistical significance. More fundamentally statistical significance is often wrongly interpreted as the probability the null hypothesis is true, given extreme results, rather than the conditional probability of obtaining extreme results if the null hypothesis is true.

Conclusions about statistical misconceptions

Our melancholy conclusion is that misconception about fundamental aspects of probability and statistics is widespread and severe, even among researchers, who presumably have completed—and taught—conventional courses in statistics. Basic intuitions are often severely and dramatically wrong, and there are damaging consequences for research and for the practical conclusions drawn from research. Teachers of statistics will probably agree with our further conclusion that it is extraordinarily difficult to help students achieve good understanding of basic statistical concepts.

We suggest two core issues are fundamental:

- misconceptions about sampling variability and the way this depends on N;
- problems with probability, especially conditional probability.

NAIVE STATISTICS: THE ANALOGY WITH NAIVE PHYSICS

Naive physics refers to intuitive beliefs people use to guide their expectations about everyday events [6].   Some naive physics beliefs are widespread, persistent, and wrong. For example many people believe a force is necessary for motion to continue, or that a ball flying around on a string will take a curved path after the string breaks. Traditional education diminishes the influence of naive beliefs, but even physics graduates often have everyday expectations which accord with naive physics.

By analogy we introduce the term naive statistics for beliefs about probability and statistics we exemplified above [1].   In statistics, as with physics, the beliefs seem to have arisen through normal life experience, although the origins of naive beliefs have been little studied. Also, in both cases the naive beliefs are used, usually unconsciously, to guide everyday expectations about the world. Because there is little research on teaching statistical concepts we look for guidance to the analogy with naive physics, and to science education generally.

LESSONS FROM SCIENCE EDUCATION AND NAIVE PHYSICS

The most general lesson is the notion of conceptual change [7]: learning will be more successful if we take account of initial naive beliefs.

Conceptual representations embodied in software
White and Frederiksen [8] addressed children's learning of electrical circuits by providing them with computer representations based on a sequence of models of increasing complexity, each corresponding to a mental model the learner should develop. The sequence of software environments allowed children to work steadily towards sophisticated quantitative concepts.

Confronting naive beliefs
Brna [9] also worked with simple electrical circuits. He designed a computer-based modelling environment ELAB in which 15-16 year olds expressed their own ideas about circuits, then explored these and confronted the inconsistencies inherent in their own naive beliefs.

White [10] described ThinkerTools, a collection of microworlds for 11-12 year olds learning elementary physics in which learners undertook modelling and game-like activities about forces and moving objects. Children formulated and tested hypotheses, and confronted any mistaken beliefs. Multiple representations were an important feature: an animation of an accelerating ball might be accompanied by symbolic representations of force and the velocity-time relation. White presented impressive evidence for the effectiveness of her approach.

Conclusions
White expressed several recommendations succinctly: 'Employ manipulable, linked representations for key abstractions. ... Make the phenomena easy to see and interpret. ... Reify the knowledge to be acquired' [10, pp. 49-50]. Our conclusions from science education are that:

- Our third initial assumption, concerning learning as conceptual change, is justified.
- The representations presented to students must be chosen carefully and must suit the target conceptualisations.
- Multiple representations yoked together offer a powerful strategy to bridge between concrete events and the symbolic representations of an academic discipline. This strategy gives a promising approach to the demanding definition of understanding expressed in our fourth initial assumption.
- Understanding must transfer to real world situations.

STATPLAY: DESIGN ISSUES

Design guidelines for educational software
We note guidelines which constitute good practice and are consistent with our initial assumptions:

1. The interface should be consistent, conform to user expectations, and be intuitive to learn and to use.
2. 'Direct manipulation' should be provided, so objects feel 'in the hand' and the user feels in control.
3. User actions should always cause some intelligible change.
4. Graphics and colour should be used for clarity and power.
5. Context-sensitive help should be provided.

Further guidelines come from education and educational computing, and from our assumptions:

6. Learners should have scope for choice and initiative; the software should feel open, and encourage exploration.
7. Interest and engagement should be maximized; game formats can be valuable.
8. Error recovery and explanation should be provided wherever possible.
9. Advice should always be available, whether on-line or in printed form, suggesting things to try next.

Science education contributes further points:

10. The representations chosen for the educational issues being targeted must be as suitable, clear and dramatic as possible.
11. Abstract symbolic representations should be used, and tied to graphic, concrete representations the user manipulates. (This is White's 'Reify the knowledge...' principle.)
12. Multiple representations are needed to integrate theoretical and practical understanding and facilitate transfer to the real world.

Goals of StatPlay
StatPlay is intended to help learners replace naive statistics beliefs with correct conceptions. It is not intended to provide a complete curriculum, but for use alongside textbooks and other software. StatPlay should be usable for lecture demonstrations, by a single learner and by a pair or small group. StatPlay must work in practice: students in practical classroom settings must benefit, and students' understanding must be practical as well as theoretical. Further, it should contribute to reform of research practice.

Software design: central issues
We discuss briefly three issues.

Design of representations of concepts: Ingenuity is needed in choosing metaphors and representations for the target concepts. (This is point 10 above.) Should we situate the learner in a casino? Or as a researcher designing experiments? Should the central metaphors be columns of numbers, pictures of distributions, or the drawing of numbered balls from a barrel? Should we work with the learner's own data, or generate samples using a statistical model? Choosing responses to such questions is the heart of the creative art that is educational software design.

Multiple representations: Points 11 and 12 specify that using multiple representations is our main strategy for achieving the understanding demanded by our fourth initial assumption. Our StatPlay experience identifies two key aspects. First, confusion and screen clutter from proliferating representations must never outweigh educational benefit and the learner's task must remain manageable. Second, multiple representations can be used in various ways:

- Simultaneous, yoked representation: as the learner drags the handle on screen the diagram transforms dynamically and the numerical values in the small windows change correspondingly. Conversely, change the value in a window and the diagram changes. The learner must be led to notice and use both (or all) representations.
- Multiple linked representation: graphical and pictorial representations are offered and the learner switches between them, noting that numerical values remain consistent, tying the two together. Perhaps the two are in different windows, usually overlaying so the learner can become familiar with each separately; later the windows are tiled for side-by-side comparison.
- Overlay representation: as in Balloon Help on the Macintosh where, if 'Symbols Mode' is selected, moving the mouse pointer to some part of a diagram could display a small overlaying window giving the corresponding statistical symbol expression, with a brief explanation.

Removing or limiting one representation can force the learner to rely on another. In this way transfer of understanding to abstract representations or to real-world applications can be encouraged.

Many simulation games, e.g. SimAnt, illustrate ingenious use of multiple linked representations, sometimes presented simultaneously, sometimes as alternatives with switching by user choice.

Learner guidance: Encouraging learner initiative and exploration must be complemented by structured activities, d guidance to make best use of exploration must be offered. Cohen *et al.* [11] gave evidence that even excellent software designed to enhance statistics understanding was not used effectively by most students unless guidance was available.

Game formats offer one attractive way to give structure: whether competition is against another learner or the clock, the learner has defined tasks to achieve. Printed worksheets, or on-line equivalents, give another model for providing structure. Even if exploration is seen as the dominant mode, a graded sequence of stored distributions, data sets, or statistical models can be provided.

STATPLAY: THE CURRENT IMPLEMENTATION

StatPlay is being developed in Visual C++ under Windows. Two microworlds, or playgrounds, have been implemented so far: the Distribution Playground allows the user to form and manipulate continuous probability distributions, to see many aspects of a distribution, and in a game format to refine intuitions about mean and standard deviation.

The Sampling Playground presents a continuous population distribution and takes random samples from this population. Various aspects including sampling distributions and confidence intervals can be examined.
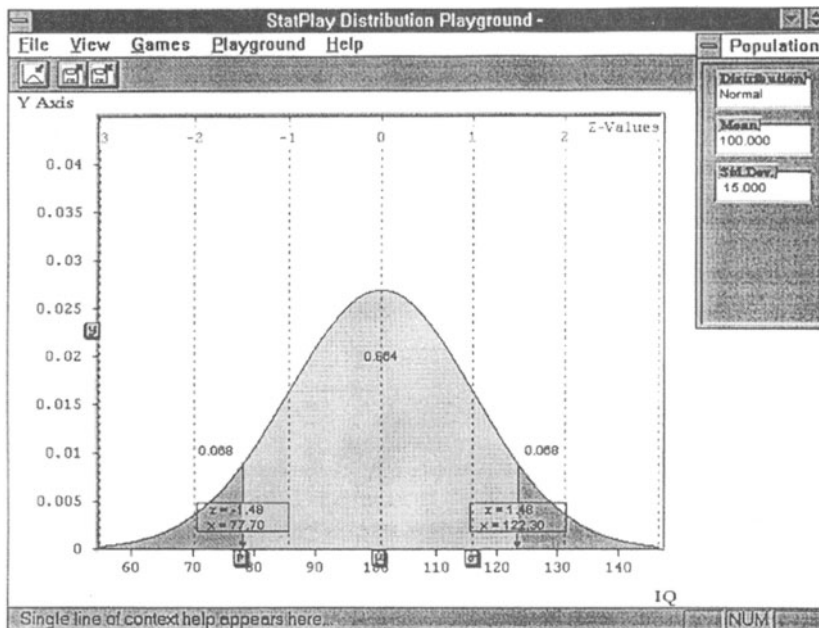

Fig. 1 The Distribution Playground

In Figure 1 the display of several aspects of the normal distribution has been turned on by the user. Dragging the handles—the small shadowed squares—changes the display dynamically. If new values of mean or standard deviation are typed into the boxes the picture transforms immediately.

In Figure 2, a skew distribution is shown in a one-person game format. The player drags the ? handles to indicate guessed values for mean and standard deviation. A digital clock, top right, indicates seconds.

In the Distribution Playground a normal (Fig. 1) or skew (Fig. 2) distribution is available, or a distribution of any shape—the freehand (Figs 3-4) —may be drawn by the user. Menu options specify aspects of a distribution to be displayed: in Fig. 1 z-score intervals are shown by vertical dashed lines, and two-tailed probabilities are shown for the coloured tail areas.
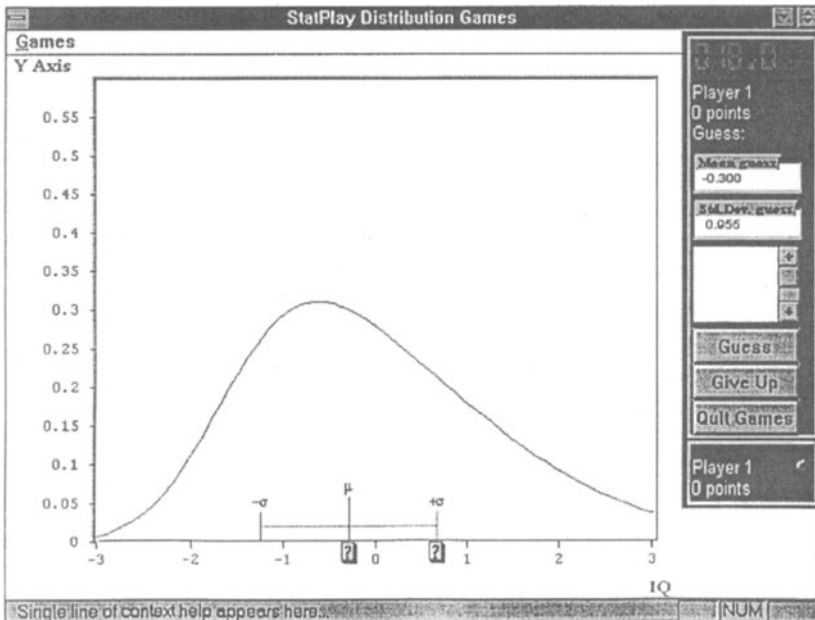

Fig. 2 A skew distribution in a one-person format

The most powerful and appealing feature cannot be shown in print: if the user clicks and drags any of the handles, the display changes immediately and smoothly. The distribution 'feels like putty under your mouse fingers'. Numerical values in the boxes at the right change dynamically also; multiple representations (curve and numerical values) are thus always coupled together.

Dragging the Y-Axis handle (near 0.02 in Fig. 1) rescales the vertical axis; dragging the mean or standard deviation handle (F or o on the IQ-Axis) transforms the whole distribution dynamically; dragging the probability handle (**p** on the X-Axis) changes the tails.

All of these features are available also for skew and freehand distributions, so users can explore the relationships among shape, probability, and positioning of the mean and z-score intervals for any continuous distribution.

Our first game formats address intuitions about mean and standard deviation. In the 1-person game (Fig. 2) you make successive guesses with feedback. In

the 2-person format you select a distribution for your opponent to guess then the roles are reversed.
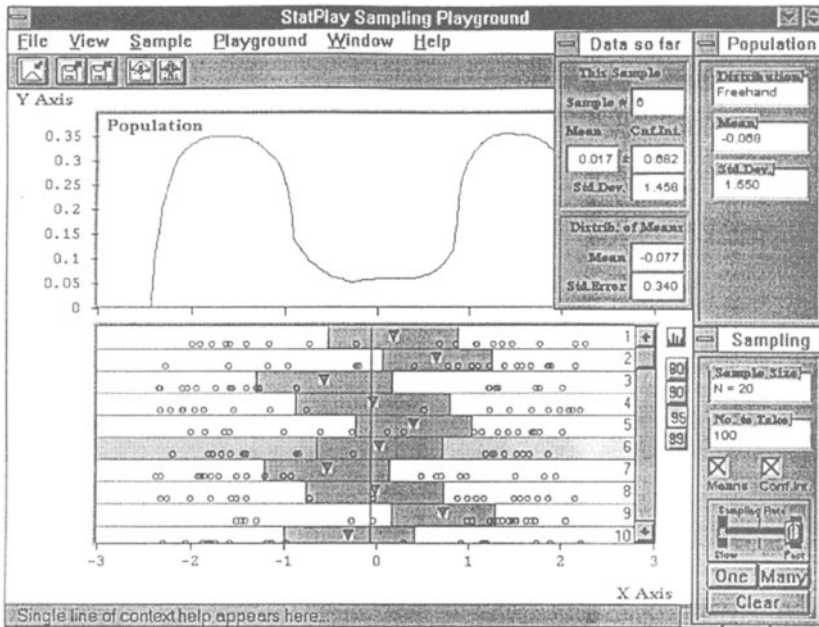

Fig. 3 A sequence of 100 samples of size $N = 20$

Figure 3 shows a sequence of 100 samples of size $N = 20$ which has been taken from a freehand bimodal distribution. Each line in the scrolling window shows one of the samples, including its mean (inverted triangle), and confidence interval (shaded bar) for the population mean (vertical line). The **95** button is pressed, indicating our choice of confidence interval.

Fig. 4 gives an alternative view of the situation of Fig. 3: the lower window shows that the frequency distribution of all 2000 individual values resembles the population distribution, while the frequency distribution of the sample means—the tiny triangles—is approximately normal with large variance as expected with N small. The central vertical line is the population mean and the light vertical lines indicate the interval expected to cover 95% of sample means.

In Figures 3 and 4 the upper part of the screen shows the population and the lower part some representation of samples taken. Three control panels give information about the population, the sampling process, and the current state

('Data so far'). The rate slider allows choices ranging from slow step-by-step sampling to rapid taking of a series of samples.
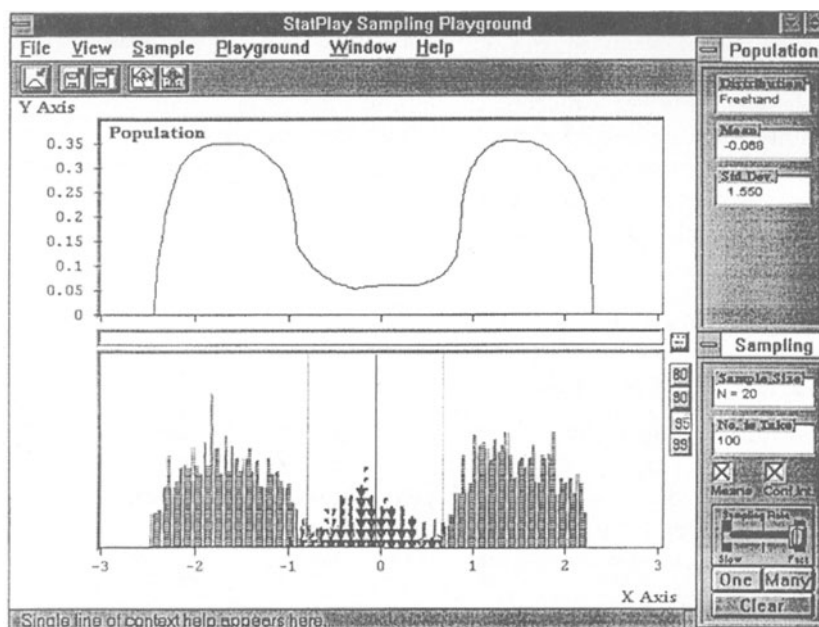

Fig. 4 An alternative view of the situation in Fig. 3

In Figure 3 the scrolling window shows the large variation in means with this small sample size ($N = 20$). Figure 4 gives an alternative view: two frequency distributions are shown—that of all the individual values and that of the 100 sample means. This sampling distribution has approximately a normal distribution, thus illustrating the Central Limit Theorem, and its standard deviation is the standard error of the mean, whose value of 0.340 is shown in the 'Data so far' control panel.

Either the scrolling window or the frequency distribution representation can be watched as it changes dynamically during repeated sampling, with the rate under control of the slider set by the user.

STATPLAY IN USE

A class of second year psychology undergraduates saw a brief lecture demonstration, then in a tutorial used StatPlay in pairs with brief guidance from worksheets. The exercise was arranged as a quasi-experiment [12], with pencil-and-paper assessment of the accuracy of several student intuitions.

One goal was for students to improve their conceptions of sampling variability, the standard error, and how these vary with N. A sensitivity index, where 1 means no sensitivity to N and 1.78 means appropriate sensitivity, increased from 1.07 before the lecture to 1.61 after the tutorial, giving initial encouragement that our approach may prove fruitful.

FUTURE DEVELOPMENTS

Planned developments include:

- Constructing a 'balloons' overlay with statistical formulas and explanation;
- Additional games;
- Playgrounds for hypothesis testing and conditional probabilities;
- On-line hints and guidance;
- Further evaluation with students.

ACKNOWLEDGEMENT

REFERENCES

1. Thomason, N. R., Cumming, G. and Zangari, M. (In press) *Understanding central concepts of statistics and experimental design in the social sciences,* in Interactive multimedia in university education: Designing for change in teaching & learning. (ed Wills, S.), North-Holland, Amsterdam.

2. Tversky, A. and Kahneman, D. (1971) *Belief in the law of small numbers.* Psychological Bulletin, **92**, pp. 105-110.

3. Schmidt, F. L. (1992) *What do data really mean?* American Psychologist, **47**, pp. 1173-1181.

4. Gigerenzer, G. (1993) *The Superego, the Ego, and the Id in statistical reasoning,* in A handbook for data analysis in the behavioral sciences: Methodological issues. (eds Keron, C. and Lewis, C.), Erlbaum, Hillsdale, NJ, pp. 311-339.

5. Oakes, M. (1986) *Statistical inference: A commentary for the social and behavioral sciences.* Wiley, New York.

6. McCloskey, M. (1983) *Intuitive physics.* Scientific American, **248**, pp. 114-122.

7. West, L. H. T. and Pines, A. L. (1985) *Cognitive structure and conceptual change.* Academic Press, Orlando, FL.

8. White, B. Y. and Frederiksen, J. R. (1987) *Qualitative models in intelligent learning environments,* in Artificial Intelligence and Education. Vol. 1: Learning environments and tutoring systems. (eds Lawler, R. W. and Yazdani, M.), Ablex, Norwood, NJ, pp. 281-305.

9. Brna, P. (1988) *Confronting misconceptions in the domain of simple electrical circuits.* Instructional Science, **17**, pp. 29-55.

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

11. Cohen, S., Chechile, R., Smith, G., Tsai, F. and Burns, G. (1994) *A method for evaluating the effectiveness of educational software.* Behavior Research Methods, Instruments, & Computers, **26**, pp. 236-241.

12. Cook, T. D. and Campbell, D. T. (1979) *Quasi-experimentation: Design & analysis issues for field settings.* Rand McNally, Chicago.