THE INSTABILITY OF YOUNG STUDENTS PROBABILITY NOTIONS

Jenni Way — University of Western Sydney
Paul Ayres — University of New South Wales

This paper draws together the findings of a series of probability studies with students aged 11 to 14 years and incorporates new data not previously reported. The collective data emphasises the fragility of probability notions and the strong effect of confirmation or refutation of outcome 'predictions' on probabilistic reasoning. Although some evidence of development of probabilistic reasoning with age was found, instruction appears to have a positive effect. The studies confirm the usefulness of videos of seemingly random experiments as research tools.

INTRODUCTION

Probability remains a pedagogically problematic topic in mathematics education at both primary and secondary levels. While the topic is recognised as being important in some countries (for example; USA, National Council of Teachers of Mathematics, 2000), there is enough concern about the efficacy of teaching the topic that it has been downgraded or even omitted in other curricula (UK, Department for Education and Employment, 1999; New South Wales Department of Education, 1998 respectively).

Much work in several avenues of inquiry is still needed:

1. Information about specific probability concepts and decision-making strategies and their relationship to other mathematical concepts and processes.
2. A coherent framework of the development of probabilistic thinking, which will support assessment procedures.
3. Development and evaluation of instruction programmes for the classroom

In recent years, work in each of these three areas has begun to overlap productively (e.g. Fischbein and Schnarch, 1997; Jones, Langrall, Thornton & Mogill, 1996, 1997, 1999; Watson, Collis & Moritz, 1997). The ultimate goal of most educational research is to inform classroom practice, and research in this third area is long overdue. However, we are far from achieving a complete model for the development of probabilistic thinking, and researchers are still discovering new information about children's perceptions of probability situations and their patterns of thinking.

Understanding of probability involves the integration of several concepts, including randomness, ratio and proportion (the structure of the sample space) and likelihood of particular events. In children, the connections between these concepts, when made, remain tenuous and can be adversely effected by occurrences that challenge their thinking (Jones, Langrall, Thornton & Mogill, 1997, 1999). For example, a child might appropriately state the most likely outcome of drawing a specified object from a box, based on either knowledge of the sample space or on a known set of previous outcomes, only to have their 'prediction' refuted by the next draw producing the least likely outcome. While the expectation that the long-term distribution of outcomes would be representative of the sample space is appropriate, believing that a small set of outcomes should be representative is inappropriate. This notion of *representativeness* has received much attention (Amir, Linchevski & Shefet, 1999; Fischbein & Schnarch, 1997; Shaughnessy, 1981), yet the quite commonly occurring scenario described above has not. The dramatic effect it can have on children's fragile probabilistic thinking is just being realised (Proposed by Truran, 1996) and confirmed by Ayres & Way, 1999, 2000).

A major hindrance to probability research with children is, ironically, randomness itself. In other areas of mathematics the concrete experiences needed to support children's learning can be prepared and predetermined. However, when engaging children in probability experiments and games, the outcomes naturally remain unpredictable. With adults and adolescents this difficulty is overcome by providing hypothetical situations and lists of outcomes, but a less abstract approach is generally desirable for children. A surprisingly little used solution is to create videos of seemingly random experiments. This enables researchers and educators to present predetermined and consistent experiences to different groups of children at different times.

This paper reports on a number of linked studies that have utilised this technology. A series of studies has been conducted by Ayres and Way (1999, 2000, 2001) with Australian students aged 11 to 14 years to investigate particular aspects of their understandings of probability. Although each of these three studies contained a new element of investigation, all of them included a common experience in the form of a pair of pre-recorded videos depicting a series of apparently random draws from a box of balls. From 1999-2001 we reported the results of these studies at various stages of completion. However, in each case, additional data was collected which was not previously reported. Consequently, the new data has enabled us to conduct further analysis, as well as more consistent statistical testing, on a larger sample size. This paper focuses on the new and consolidated findings.

Video recordings as a research instrument

The major research tools used in the studies were video-recordings. Two videos were made which featured a presenter making thirty selections of coloured balls from a box, with replacement. In each video, 19 whites (63% of the total selections), 7 blues (23%) and 4 yellows (13%) were drawn from the box. For each group of five selections, the total number of whites, blues and yellows balls that occurred were identical for both videos. Consequently after each set of five selections the progressive experimental probabilities of each colour occurring were identical. However, for the 6^{th}, 11^{th}, 16^{th}, 21^{st} and 26^{th} selections, the colours varied according to the video. For one video, which we called the *typical sequence video*, a white was selected in four of these five positions. For the second video, called the *non-typical sequence video*, one of the less likely colours (blue or yellow) occurred in place of a white four times out of five. As students were required to predict the most-likely colour for the next draw after observing the five previous selections, these differences were highly significant. Students who consistently chose the most frequently occurring colour (white) would be successful in their predictions if they viewed the *typical sequence*, but unsuccessful if they viewed the *non-typical sequence.* As a result of this design we were able to investigate how confirmation or refutation of success affected student choice.

Although the outcomes of these videos were manipulated and not random, it was expected that students observing these videos would believe that they were genuine random experiments. This was tested (see Ayres & Way, 1998a, 1998b) and found to be the case. Furthermore, over the whole series of experiments completed so far, no evidence has ever emerged that students doubted the validity of the videos. By controlling the colour sequences we were able to not only compare how students reacted differently to the two videos, but also see how different groups of students reacted to the same videos. Consequently, it was possible to make meaningful comparisons across studies, as the instruments are identical. A major empirical difficulty in conducting probability experiments of this nature is that random generators produce random outcomes. Video-recording techniques, such as the one described here, can overcome this problem.

RESULTS

Prediction analysis

For each study, students observed one of the videos and were asked to make six predictions. Students were told that the prediction tasks were a game and they should try to predict as many correct colours as possible. Because the selection of white (the most frequent colour) was considered an important statistic, as it indicated probabilistic reasoning, the number of whites chosen was reported in each study. Furthermore, as students often changed their strategies over the course of the trial, prediction trends were noted by recording the number of whites chosen in the first and last three predictions. It should be noted that none of the participants in these studies had any formal classes in understanding likelihood, as probability was not part of the syllabus for these students

1999 Study

The main goal of the 1999 study was to investigate how students were influenced in their probability judgements by confirmation or refutation of their predictions. Students from the same school participated from two grades (Grade 6, aged 12 and Grade 7, aged 13). Half the student from each grade observed one of the two videos. The mean number of white balls by group and location selected are reported in Table 1. The inclusion of the Grade 7 data, previously not reported, doubles the sample size and allows a 2x2 (grade x typicality) ANOVA to be conducted on the total number of white balls selected. A significant main effect for typicality was found; $F(1,116) = 4.3$, $p < 0.05$; indicating that the students who viewed the *typical* video selected white more times than those who viewed the *non-typical* video. No main effect was found for grade: $F(1,116) = 2.5$, $p = 0.12$.

Table 1: Mean number of white balls selected by group and location (1999 study)

	Typical sequence			Non-typical sequence		
	1st3	2nd3	**Tot**	1st3	2nd3	**Tot.**
Grade 6 N= 60	1.1	1.7	**2.8**	1.2	1.0	**2.2**
Grade 7* N=60	1.3	1.7	**3.0**	1.5	1.2	**2.7**
Combined	1.2	1.7	**2.9**	1.4	1.0	**2.5**

*Data not previously reported

A notable feature of this study was how prediction strategies changed over the course of the task. For the *typical* group (Grades 6 and 7 combined), students chose significantly more whites in the second half of the task (final three predictions) than on the first (first three predictions) under a paired t-test: $t\ (61) = 3.6$, $p < 0.01$. In contrast, the *non-typical* group chose significantly less white balls for the second half compared with the first half under the same test: $t\ (57) = 3.3$, $p < 0.01$. These comparisons indicated that students who were being rewarded for choosing the most likely colour (*typical* group) continued more frequently with this strategy, whereas students who were not being rewarded (*non-typical* group), reduced their selections of the most likely colour.

2000 Study

In the 1999 study, students were not told about the sample space (ratio of the coloured balls in the box) prior to starting the trial. However, in 2000, the effect of knowing or not-knowing the sample space was investigated. The experimental design of the previous study was extended by subdividing further each of two groups who received either a *typical* or *non-typical* video into two subgroups, one of which knew the sample space and the other did not. In addition to the previously reported data on Grade 8 students, a cohort of Grade 7 students participated from the same school. This data is reported in Table 2. A 2 x 2 x 2 ANOVA (grade x typicality x sample space) was conducted. A significant main effect was found for grade: $F\ (1, 100) = 15.5$, $p < 0.01$; indicating that the older group of students selected more whites than the younger group. There was also a significant main effect for typicality: $F\ (1, 100) = 12.6$, $p < 0.01$; indicating that students who observed the *typical* video predicted more whites than those who viewed the *non-typical* video. However, there was no main effect for sample space: $F\ (1, 100) = 0.2$, $p = 0.64$; indicating that knowing or not-knowing the sample space made no difference to student predictions. No significant interactions were found.

Analysis of the prediction trends revealed that students who viewed the *typical* video had no significant difference in their choice of whites between the first and second three predictions under a paired t-test: $t\ (53) = 0.2$, $p = 0.82$. However, students who viewed the *non-typical* video predicted significantly less whites during the second half of their predictions than the first half: t (53) = 3.5, p < 0.01.

Table 2: Mean number of white balls selected (2000 study)

	Typical sequence						**Non-typical sequence**					
	Unknown Sample Space			Known Sample Space			Unknown Sample Space			Known Sample Space		
	1st3	*2nd3*	***Tot***	*1st3*	*2nd3*	***Tot***	*1st3*	*2nd3*	***Tot***	*1st3*	*2nd3*	***Tot***
Grade 7* (n=48)	2.0	1.6	**3.6**	1.9	1.5	**3.4**	1.5	1.1	**2.6**	1.3	0.8	**2.1**
Grade 8 (n=60)	1.8	2.2	**4.0**	1.8	2.2	**4.0**	1.8	1.1	**2.9**	2.2	1.7	**3.9**
Comb.	1.9	1.9	**3.8**	1.8	1.9	**3.7**	1.7	1.1	**2.8**	1.8	1.3	**3.1**

*Data not previously reported

Interview data

Twenty-four (three chosen at random from each cell in the study) of the 108 participating students were asked to give reasons for each prediction made. The four most frequent reasons given were: a) more white balls in the box (42%), b) the colours formed a pattern (22%), c) guessed (15%), and d) a particular colours turn to come up (15%). Students who stated that they chose white because there were more whites in the box were demonstrating probabilistic reasoning. However, the frequency of this response varied according to which video was viewed. By recording group means for the number of times this "more whites" reason was given a 2 x 2 (typicality x sample space) ANOVA could be conducted (grades combined). There was a significant main effect for typicality: F (1, 20) = 6.5, p < 0.05; indicating that students who viewed the *typical* video gave more reasons consistent with probabilistic reasoning than those who viewed the *non-typical* video. No significant main effect was found on knowing the sample space or not. By further analysing the "more whites" reason over the first and second three predictions made, a further difference was found. Students who viewed the *typical* video, gave the "more whites" reason significantly more times during the second three predictions (2.1) than the first three (1.1): t (11) = 3.4, p < 0.1; indicating that they became more convinced as the trial progressed. However, for those who viewed the *non-typical* video, no significant difference was found between the two phases: t (11) = 1.1, p =0.31. Although, "more whites" was offered as the reason more on the first 3 predictions (mean = 1.1) than the second 3 (mean = 0.8). Overall, the trends for the reasons given, consistently matched the predictions made.

2001 Study

The main aim of this study was to investigate the effect of instruction (small-group practical activities) on predictions. As reported previously (Ayres & Way, 2001), the different instructional experiences and group dynamics affected student predictions when individuals responded to the *non-typical* video. In addition to this data, students were asked to give reasons why they made each prediction. The mean number of times the "more whites" response was given is recorded in Table 3. Although, knowing or not-knowing the sample space failed to reach significance, knowing the sample space produced higher means for both the number of whites chosen and the "more whites" reason, and suggested there may be a real effect for a larger sample. Furthermore, consistent with the other studies, there was a significant decrease in the number of whites chosen; $t(21) = 2.7$, $p < 0.05$, and the "more whites" reason; $t(21) = 3.5$, $p < 0.01$; from the first 3 predictions to the last 3.

Table 3: Mean group responses to the untypical video (2001)

Grade 6 students (n = 24)	Unknown Sample Space			Known Sample Space		
	$1^{st}3$	$2^{nd}3$	**Tot**	$1^{st}3$	$2^{nd}3$	**Tot**
Number of whites chosen	2.3	1.6	**3.9**	2.6	1.9	**4.5**
Number of times "more whites" given as a reason*	2.2	1.3	**3.5**	2.6	1.6	**4.2**

*Data not previously reported

CONCLUSIONS

The inclusion of new data and its analysis has strengthened and extended the findings of the 1999-2001 studies. Overall, the findings fall into three main categories. Firstly, it is clear that students in this age range appear to develop an understanding of likelihood with age, even without instruction. By choosing a high percentage of the most-likely colour throughout these studies, students have demonstrated some understanding of basic probability. Interview data, collected on students' reasons for particular predictions, have reinforced this conclusion. However, many misconceptions have also been observed. Notably, many students believe that colours that have not occurred for a while are then more likely to occur- "it's that colour's turn to come up". This *negative recency* effect has been observed in other studies (see Fischbein and Schnarch, 1997). Of particular interest has been the discovery that students often look for patterns. This finding, which is easily explained considering the emphasis placed on patterns in modern curricula, may make understanding randomness for students of this age, more difficult, especially if outcomes are displayed in pattern-like sequences found in these studies.

Secondly, students in this age group have a fragile commitment to using probabilistic reasoning consistently. The differing responses to the two videos have demonstrated this effect conclusively. Students, who viewed the *non-typical* video, switched from choosing the most-likely colour into strategies based around various misconceptions. In contrast, students who viewed the *typical* video tended to choose the more-likely colour more often, as the trial progressed. Refutation or confirmation of 'prediction' is an important factor in this type of task.

Thirdly, instruction appears to have a positive effect. Students presented with the *non-typical* video after instruction, extensively chose the most frequently colour and argued that they did so because it was more likely. Nevertheless, this strategy still decreased for many students as a result of their predictions being refuted. However, of particular note, was the high number of whites chosen by the Grade 6 students compared with all other students in these studies, many of whom were older. In particular, students who were aware of the sample space had a mean score of 4.5 out of 6 - the highest score for any group in all three studies. Although, knowing the sample space or not, did not produce a significant difference when directly tested, it was done so only on students who received no instruction. It is feasible that following instruction, knowledge of the sample space may become more meaningful.

Further research directions follow directly from these studies. The previous studies were deliberately limited to one type of random generator, a particular sample space structure and two specified sequences of outcomes. However, future research could expand to include a wider variety of random generators, such as dice and spinners; a range of sample space ratios; and various sets of frequency data. The value of using video-recorded probability trials in such research has been established. Therefore, to enable the use of videos as instructional aids and/or assessment tools a range of videos could to be developed. Using such videos, the effectiveness of instruction can be directly tested, as well as the types of instruction, and the influence of knowledge of the sample space.

Finally, in reference to the debate about inclusion of *chance* into school curriculum, it should be acknowledged that it is a difficult concept for young children to understand; nevertheless, our research suggests that students as young as eleven can benefit from instruction.

REFERENCES

Amir, G., Linchevski,L. & Shefet, M. (1999). The probabilistic thinking of 11-12 year old children. In Zaslavsky, O., (Ed.) *Proceedings of the Twenty Third Conference of the International Group for the Psychology of Mathematics Education,* Haifa, Israel, (2) 25-32.

Ayres, P & Way, J (1998a). The effectiveness of using a video-recorder in randomly probability experiments. In Pereira-Mendoza, L., Kea, L., Tang, W. & Wong, W (Eds). *Published Proceedings of the Fifth International Conference on Teaching Statistics* (ICOTS-5). Vol 3, p 1415-1421. Singapore, June 1998.

Ayres, P & Way, J (1998b). Factors influencing predictions made on randomly generated sequences. In Kanes, C., Goos, M. & Warren, E. (Eds.). *Published Proceedings of the Mathematical Education Research Group of Australasia (MERGA 21),* Gold Coast, Australia, 74 - 81.

Ayres, P & Way, J (1999). Decision-making strategies in probability experiments: the influence of prediction confirmation. In Zaslavsky, O., (Ed.) *Proceedings of the*

Twenty-third Conference of the International Group for the Psychology of Mathematics Education. Haifa, Israel, (2) 41-48.
Ayres, P, & Way, J. (2000). Decision-making strategies in probability experiments: the influence of prediction confirmation *Published proceedings of the 24th Psychology of Mathematics Education Conference*, Hiroshima, Japan, (2) 33-40.
Ayres, P & Way, J (2001). The effects of instruction on likelihood misconceptions. In van den Heuvel-Panhuizen, M, (Ed.) *Proceedings of the Twenty-fifth Conference of the International Group for the Psychology of Mathematics Education.* Utrecht, The Netherlands, (2) 73-81.
Department for Education and Employment (UK) (1999). *The National Numeracy Strategy: Framework for Teaching Mathematics from Reception to Year 6*. DfEE: London
Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal of Research in Mathematics Education, 28*: 1, 96-105.
Jones, G., Thornton, C., Langrall, C. & Mogill, T. (1996). Using probabilistic thinking to inform instruction. In L, Puig & A, Gutierrez (Eds.), *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education,* Valencia, Spain, (3) 137-144.
Jones, G., Langrall, C., Thornton, C. & Mogill, T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32,* 101-125.
Jones, G., Langrall, C., Thornton, C. & Mogill, T. (1999). Using probabilistic thinking to inform instruction. *Journal for Research in Mathematics Education, 30(5),* 487-519.
National Council of Teachers of Mathematics (USA) (2000). *Principles and Standards for School Mathematics*. Reston, Va.: NCTM..
New South Wales Department of Education (1998). Mathematics K - 6. NSW Department of Education:Sydney
Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic errors and decisions. In A. Shulte & J. Smart, *1981 Yearbook: Teaching Statistics and Probability*. (pp.90-100.), Reston, Va.: NCTM.
Truran, J. (1996). Children's misconceptions about the independence of random generators. In L, Puig & A, Gutierrez (Eds.), *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education,* Valencia, Spain, (4) 331-338.
Watson, J. Collis, K. & Moritz, J. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9 (1),* 60-82.