FACTORS INFLUENCING PREDICTIONS ABOUT RANDOMLY GENERATED SEQUENCES

Paul Ayres & Jenni Way
UWS Nepean
Australia

The detailed analysis of four probability experiments conducted with Grade 5 and 6 students revealed trends and patterns in both the group and individual data. These results suggested that certain variables in the experiments, such as particular sequences of outcomes and the confirmation/refutation of student predictions, influenced the students' decision making strategies. The use of video recordings of deliberately controlled probability experiments offers the potential to systematically explore these influential factors with large samples of students.

Introduction

Several research studies exploring probabilistic reasoning have highlighted the influence of particular sequences of randomly generated outcomes on decision making strategies (for example: Peard, 1995; Shaughnessy, 1981; Tversky & Kahnemann, 1982). One such strategy is *representativeness* (Kahneman & Tversky, 1972), which is the expectation that a random set of outcomes should be representative of the composition of the known sample space. Related to representativeness is the type of thinking known as *negative recency* or *gambler's fallacy* , where there exists the expectation that as the frequency of a particular outcome increases the probability of that outcome ocurring decreases. For example; when repeatedly flipping a coin, a run of heads would lead to the expectation of the next flip being a tail. The opposite strategy, in this case of predicting a head because that's the trend, is referred to as *positive recency*.

Much of the research into misconceptions and inappropriate reasoning strategies has been with adults, using written tasks or 'tests', in which preconceived sets of outcomes have been presented to the subjects. However, most researchers working with children prefer to use real random generators to accommodate children's need for concrete experiences (for example: Carlson, 1970; Fischbein, 1975; Hoemann & Ross, 1971; Truran, 1992; Way, 1996). Random generators naturally produce sequences which vary from trial to trial. Consequently the ability to test large groups of students with controlled variables is restricted. The use of video-recorded probability experiments can overcome these research difficulties by providing both a realistic medium for children, and controlled sequences of outcomes (see Ayres & Way, 1998).

Truran (1996) analysed the nature of primary and secondary students's responses to a task in which they were required to state the most likely outcome of each of nine draws (with replacement) from a sample space of three green and one blue. The use of the negative recency heuristic in such tasks has been quite well documented, but more interestingly, Truran examined the effect that either confirmation or refutation of the predicted outcome had on the following prediction. In other words, he analysed the changes in prediction in regards to the next outcome. One finding was that when green (the more likely outcome in this experiement) was predicted, it didn't really matter whether the next draw confirmed or refuted that prediction. However, if the less likely colour (blue) was predicted, the subject was highly likely to change the prediction, particularly if the following outcome refuted the blue prediction. The strong influence that knowledge of the sample space had on the students' predictions was obvious, so this raises the question of what effect the absence of this knowledge might have on prediction patterns.

Probability experiments with hidden sample spaces force the subjects to rely on the information provided by the experimental outcomes (relative frequencies) when considering the likelihood of various outcomes. Fischbein (1975) established that even young children possess the *intuition of relative frequencies* and can therefore estimate the composition of an unknown sample space given adequate frequency data. But how much data is adequate, and how representative of the actual sample space does this frequency data have to be to stimulate reasonably accurate estimates?

The two main research questions currently of interest to the authors are as follows: a) What is the relationship between the degree of representativeness of frequency data produced from an unknown sample space and students' probability judgements? and; b) What is the relationship between the confirmation or refutation of student predictions and students' decision making strategies?

As part of an on-going project to investigate the development of probabilistic reasoning the researchers have completed two preliminary studies. Although each study had a different specific purpose, the design of the experiment and the type of data collected in both studies were similar. This particular paper reports the patterns and insights which have emerged by bringing the findings of the two studies together and applying for the first time, common data-analysis techniques. Firstly, brief summaries of the studies are given.

Study 1

The first study (Ayres, 1996) was exploratory and investigated Grade 6 responses to a sequence of coloured balls drawn at random from a bag containing six orange, three white and one yellow, with the actual contents of the box unknown to the students. (Note that these colours have been altered from the original study to standardise the reporting of both studies). Students were required to specify the colour they thought most likely to be drawn after every five selections. An introductory activity and discussion was used to assure the students that the researchers did not expect them to actually *know* the outcome of the next draw. A record of the outcome of each selection was displayed for the class. Students made six predictions in total and their individual choices are shown in the Appendix. Two classes (referred as 1A and 1B) participated in this study and observed radically different sequences (consistent with naturally occurring random outcomes), and consequently made differing predictions from a class perspective. The colour sequences that groups 1A and 1B observed contained a total of 73% and 53% oranges respectively. Although both sequences produced experimental data consistent with orange being the most likely colour to occur, the sequence observed by 1A contained a greater number of orange than the expected value (60%). The students in the group 1A constantly predicted the colour orange (93% of the time). It was therefore concluded that the students in this group were guided by experimental outcomes and were able to link ratio with likelihood. In contrast, the students in 1B only choose orange 43% of the time, suggesting that the different frequency data influenced their predictions.

Study 2

Following this study the researchers concluded that to test large groups of students in this domain, and understand the reasons for their decision-making, it was desirable to control the variables. As a result the second study (Ayres & Way, 1998) tested the effectiveness of using a video-recording as a method of controlling variables and reproducing random sequences. In this study, a group of Grade 5/6 students were shown a video-recording of balls being drawn from a box and asked to specify the most likely result of a draw in a similar fashion to the first study. However, the video-recording was a reproduction of a sequence of random colours which a parallel class of Grade 5/6 students had observed *live* and made predictions on. Quantitative and qualitative data revealed that the group of students who observed the *video* responded to the prediction tasks in a similar fashion to the live group and were not concerned about its legitimacy: students believed that the sequence was random, and not artificial.

Two Study Analysis

Overall, the two studies produced three sets of frequency data which are recorded in Table 1. Given the nature of the prediction tasks in these studies, it can be assumed that students who were guided by experimental probability were overall more likely to predict the most frequently occurring colour. Over the six trials the mean number of oranges predicted was 5.6 (SD = 0.9) for Group 1A and 2.6 (SD = 1.4) for Group 1B. Similarly the mean number of oranges predicted was 2.9 (SD = 1.33) for the Live Group and 2.5 (SD = 1.4) for the Video Group. Clearly, only Group 1A adopted a consistent policy of predicting the most frequently occurring colour. In contrast, the other three groups all predicted the most frequent occurring colour less than 50% of the time. Although, there appears to be little difference in overall means for these three groups, substantial differences can be found in the prediction patterns.

Table 1: The random colour sequences produced in the two studies

Groups	*Colour Patterns*
1A	OWOOO WOOOO OYOOO OOWOO OOOWW OWOOW W
1B	WOOYW WYOOO YOOOO WWOOO OOOYW WOWYW W
Live and Video	OYWYO WOWOO WOWOW OOOOO YOOOO WOOOO

Notes: a)The underlined colours indicate the outcome immediately following each request for a prediction; b) O = orange, W = white, Y = yellow.

Prediction profiles for each group were found by calculating the frequency of each colour selected at every prediction point (see Table 2). Whereas, Group 1A was very consistent at each prediction, the other groups differed considerably, especially on their first three selections. Although these differences could be due to random prediction factors between groups, there are also intrinsic differences contained within the sequences themselves which may have been influential.

Table 2: Colour selections (%) at each prediction point

	Group 1A			Group 1B			Group Live			Group Video		
Predictions	O	W	Y	O	W	Y	O	W	Y	O	W	Y
First	95	5	0	45	10	40*	17	47	37	61	26	13
Second	90	10	0	30	40	30	47	23	30	4	30	65
Third	85	0	15	55	35	10	43	33	23	26	39	35
Fourth	95	5	0	60	5	35	57	20	23	57	22	22
Fifth	95	5	0	40	25	35	67	3	30	30	52	17
Sixth	95	5	0	30	50	20	60	13	27	65	17	17
Means	93	5	3	43	28	28	49	23	28	42	31	28

* Note: other colour selected

Group Trends

By calculating the unfolding experimental probabilities (see Table 3) the differences between sequences can be observed. For example, for the Live and Video groups the first four colour outcomes produced a high (0.5) experimental probability of yellow occurring (2 yellows out of four). Consequently the selection of orange was not necessarily an informed choice by these groups at this stage. However, by the 20th selection the experimental colour proportions were more representative of the actual theoretical probabilities for both the Live & Video groups; and similarly for Group 1B.

Table 3: Cumulative experimental probabilities (%) per cluster for each random sequence.

	Group 1A	Group 1B	Groups Live & Video
Clusters	O W Y	O W Y	O W Y
First	80 20 0	40 40 20	25 25 50
Second	80 20 0	50 30 20	44 33 22
Third	80 13 7	60 20 20	50 36 14
Fourth	80 15 5	60 25 15	58 32 11
Fifth	76 20 4	60 24 16	63 25 13
Sixth	73 23 3	53 30 17	66 24 10

(Note: Each experimental probability percentage has been calculated by combining the results of each prediction cluster. For example, the first experimental probabilities for groups 1A and 1B are calculated after five selections; the second after ten selections and so on).

In the Ayres & Way (1998) study the slow development of representative experimental probabilities were taken into account by comparing the first three predictions with the last three. It was argued that this measure would be a stronger indicator of probabilistic reasoning as the experimental probabilities were converging more towards the theoretical in the latter half of the sequence. By calculating the mean number of oranges (most likely colour) for the first and last three predictions it was possible to assess whether the groups had refined their strategies for the last three predictions. This data is reproduced in Table 4, along with a similar analysis for Groups 1A and 1B. It should be noted that this analysis was not previously applied for Groups 1A and 1B. One tailed (improvement expected) Wilcoxon signed-rank tests for matched pairs were conducted. As recorded in the Ayres & Way (1998) study both the Live & Video groups demonstrated a significant increase in the number of oranges selected, as the frequency data became more representative of the theoretical probabilities, and hence the sample space.

Table 4: Mean number of predictions made of the most frequently occurring colours.

	1A	1B	Live	Video
Mean number of "most likely" colour chosen on first 3 predictions	2.7	1.3	1.1	0.9
Mean number of "most likely" colour chosen on last 3 predictions	2.9	1.3	1.8	1.6
One-tailed Wilcoxon signed-rank tests for matched pairs	T (3) = 0	T(16) = 68.5	T(21) = 36*	T(19) = 37.5*

* Significant at the 99% level

Significance was not reached by Group 1A because n=3 is too small a number to evaluate under the Wilcoxon test where n=5 is usually required; although group means have increased. It should be noted that as 75% of this group selected all oranges anyway, there is little capacity for more oranges to be chosen. Only Group 1B did not show any indication of an overall group movement towards favouring the most likely colour. Individual analysis indicates that 8 students in Group 1B increased their choice of orange; whereas eight decreased. Only one student (5% of the group) completed a sequence of three or more oranges to finish the prediction sequence. This contrasts with five (21%) students in the Video group and 10 (33%) in the Live group who finished with 3 or more oranges. Further more, four students in Group 1B predicted three or more oranges to begin with, but then changed to another colour.

Factors which may have influenced Group 1B on this measure can be found by analysing the colour sequences further. Table 3 indicates that the experimental frequencies quickly converge towards the expected values by the third prediction. Although the experimental probability of orange occurring does decline from 60% to 53% by the last prediction, it is still the most frequently occurring colour, and arguably does not explain the low means prediction rate of 1.3 for the last three predictions. Ayres (1996) suggested that the lack of success in their predictions by this group influenced their strategies. More evidence can be found to support this argument by examining the colours which did occur immediately after the prediction points. (These colour outcomes form a subset of the frequency data and are underlined in Table 1). For Group 1A and the Live & Video groups the most frequently occurring colour (orange) was drawn four times out of five immediately after each of the predictions. (For this analysis the sixth prediction need not be considered as students were not required to make further predictions following its selection). In contrast, for Group 1B, an orange only occurred once after each of the five predictions. This may well be a significant factor. Students in Groups 1A, Live and Video applying a positive recency strategy (in regards to the subset of post-prediction outcomes) had their predictions confirmed, but students in Group 1B applying this strategy did not. Analysis of prediction success rates indicated that Group 1B had the lowest success rate of 1.5 (SD = 1.1) compared with 3.7 (SD = 0.64) for Group 1A, 2.2 (SD = 1.08) for the Live group and 1.9 (SD = 1.1) for the Video group. This finding suggests another variable which may need to controlled in future experiments.

Individual Analysis

Further insights can be gained by examining individual responses in more detail. In group 1A, orange was so dominant in the frequency data that very little additional information can be extracted, except at the third prediction, where three students chose yellow. This was the only time that yellow was chosen and followed a yellow occurring in the previous cluster which may have influenced these students' choices. There is some evidence to suggest that some students may have been utilising a positive recency strategy in some predictions. For example, following a white occurring three times in the sixth cluster, 50% of the students in Group 1B selected a white at the next prediction point. In addition, following two whites occurring in the third cluster, nine students (39%, the highest number) in the Video Group chose white, and ten students (33%) in the Live group. In contrast, there were examples of colours chosen where students may have adopted a negative recency strategy. Fifteen students (65%) chose yellow for their second prediction following the non-selection of a yellow in the previous cluster. Similarly 12 students (52%) chose white for their fifth prediction in the Video Group, following the non-occurrence of a white for nine selections. It is interesting to note that five students (two in Group 1B, two in the Live group and one in the Video Group) did not chose the most frequently occurring colour at all. In addition, it can be seen from the means in Table 2 that yellow was chosen as frequently as white, despite having the least likelihood of occurring (see Table 3).

Changes in Prediction

Further analysis (conducted for the first time on the data from both studies) of individual responses revealed that many students changed the choice of their colours frequently. The extent of these changes was calculated by counting each change of colour from one prediction to the next, thus giving a maximum of five possible changes. For example, Student 1 in Group 1B had a prediction sequence of O W W O Y W giving a total of four changes. The mean number of changes by group are shown in Table 5.

To investigate the extent to which changes are influenced by successful predictions, changes were categorised according to whether they followed confirmation or refutation of predictions. For each student the number of changes made following either outcome was calculated as a proportion. For example, if a student made four correct predictions which were followed by two changes in colour following each success, then the student had a ratio of 0.5 changes per successful predictions. Group means for this data are also recorded in Table 5.

Table 5: Changes in colour predictions

	1A	1B	Live	Video
Mean Number of Colour Changes	0.8 (SD = 1.4)	3.8 (SD = 1.3)	2.7 (SD = 1.4)	3.3 (SD = 1.0)
Mean number of proportional changes following a *successful* prediction	0.12 (SD = 0.3)	0.7 (SD = 0.4)	0.4 (SD = 0.3)	0.7 (SD = 0.3)
Mean number of proportional changes following *unsuccessful* prediction	0.23 (SD = 0.4)	0.8 (SD = 0.3)	0.7 (SD = 0.3)	0.6 (SD = 0.4)

As might be expected from a group who constantly favoured one colour, Group 1A demonstrated the least number of changes with a mean rate of 0.8 changes per five selections. Furthermore, it can be seen that this measure appears to form an inverse relationship with the prediction success rates of the groups. As reported above, the mean number of correct predictions made over the first selections were 3.7, 2.2, 1.9 and 1.5 respectively for groups 1A, Live, Video and 1B, which may indicate a negative correlation with the corresponding number of changes made (0.8, 2.7, 3.3 and 3.8). In other words, students who successfully predict a colour are less likely to change that prediction at the next selection point. Further evidence can be found to support this possibility by examining the changes made following confirmed or refuted predictions. The two most successful groups (1A & Live) at predicting, had the lowest rates of changes (see Table 5) following successful predictions. But these rates increased considerably (1A: from 0.12 to 0.23; Live: from 0.4 to 0.7) following an unsuccessful prediction. Although Group 1B followed this trend to a lesser degree (from 0.7 to 0.8), the Video group did not (from 0.7 to 0.6); a point which requires further investigation.

Conclusions

Although, a certain amount of caution should be shown as a result of examining the combined data from two independent studies, more patterns and trends emerged than could be identified from the studies individually. In any domain involving random generators, different sets of outcomes emerge which make the analysis of the data complex. However, by combining the data from four different groups, and searching for similarities, contrasts and patterns, the researchers were able to identify factors that should be considered when designing further investigations. The video-recording technique allows control over the frequency data, which can then be presented to many groups of children. The relationships between variables in the frequency data and the students' predictions which have been tentatively identified, can now be used to inform the ways in which the researchers manipulate these variables in future studies. As suggested by other researchers (for example: Truran, 1996; Way, 1996), the strategies employed by children in probabilistic situations are more complex than can be explained by currently identified heuristics. Although still in its early stages of development, the overall methodology outlined in this paper, has the potential to unravel some of these complexities, and more clearly identify some aspects of the development of probabilistic reasoning.

Note: The researchers wish to acknowledge the contribution made by the intial reviewers of this paper. This research was supported by a UWS Nepean Research Grant.

References

Ayres, P. (1996). An investigation into Grade-6 responses to a random generator. *Australian Association for Research in Education and Educational Research Association (Singapore) Joint Conference*, Singapore, November, 1996. (http://www.swin.edu.au/aare/conf96p.htm, AYREP96.213).

Ayres, P. & Way, J. (1998) The effectiveness of using video-recording to reproduce randomly generated sequences in probability research. Paper accepted for presentation and publication at *The Fifth International Conference on Teaching Statistics*. Singapore. June, 1998.

Carlson, J. (1970). The development of probabilistic thinking in children: A comparison of two methods of assessment. *Journal of Genetic Psychology, 116,* 263-269.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children.* Dordrecht: Reidel.

Hoemann, H., & Ross, B. (1971). Children's understanding of probability concepts. *Child Development, 42,* 221-236.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3,* 430-454.

Peard, R. F. (1995). Student decision making in a game of chance and misconceptions in probabilistic reasoning. In Atweh, B. & Flavel, S. (Eds.) *Galtha: Proceedings of the Eighteenth Annual Conference of the Mathematics Education Research Group of Australasia.* (pp. 440-445). Darwin, Australia.

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, Vi.: National Council of Teachers of Mathematics.

Truran, J. (1992). Children's understanding of the independence of random generators. In B. Southwell, B. Perry & K. Owens (Eds.) *Space - The First and Final Frontier: Proceedings of the Fifteenth Annual Conference of the Mathematics Education Research Group of Australasia.* (pp.541-550). Richmond, Australia.

Truran, J. (1996). Children's misconceptions about the independence of random generators. In L, Puig & A, Guitierrez (Eds.), *Proceedings of the Twentieth International Conference for the Psychology of Mathematics Education.* Vol. 4.(pp. 331-338) Valencia, Spain.

Tversky, A. and Kahneman, D. (1982). Judgement under uncertainty: Heuristics and biases. In Kahneman, D., Slovic, P. & Tversky, A. (Eds.), *Judgement Under Uncertainty: Heuristics and Biases.* (pp. 3-22). Cambridge: Cambridge University Press.

Way, J. (1996). Children's Strategies for Comparing Two Types of Random Generators. In L. Puig and A. Gutierrez (Eds.) *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education.* Valencia, Spain. (pp. 419-426).

Appendix: Predictions made by each student in both Studies

	Study 1			*Study 2*	
	Group 1A	*Group 1B*		*Live Group*	*Video Group*
1	OOYOOO	OWWOYW		OOWOYW	YWOYOO
2	OOOOOO	WYOYWY		OWYOYY	WYWOOO
3	OOOOOO	YWOYWY		YOOOOO	WWYOOW
4	WOYOOW	WYWYOW		OYYOYY	OYOOOO
5	OOOOOO	OOOYOO		YOOWOO	YYWOWO
6	OOOOOO	OYWOWY		YOOOOO	YYOWWY
7	OOOOOO	OOOOYW		YOWOOO	OWWOWO
8	OWOOOO	OYOWYO		WWWWOO	OYWYWW
9	OOOOOO	OOOOWO		YWOYOO	OYYWWY
10	OOOOOO	OWWOWO		WYWOYO	WYWWWO
11	OOOOOO	YYWYYY		YYYYYY	OYYYWO
12	OOOOOO	OWOOWW		WYOYOW	OWWOYO
13	OOOOOO	YWOOOW		WWOOOO	WYYYWY
14	OOOOOO	YOWOYO		OYYWOY	OYWOYO
15	OOYOOO	YOYOOW		WYOWOW	OYOOOO
16	OOOOOO	YYWYOW		YYYOOO	WOYWWW
17	OWOWOO	YYWYOW		YYWWOO	OYOOOO
18	OOYOWO	OOOOOW		WOWYYY	OYWOYW
19	OOOOOO	YWYOYO		YOOWOO	OYYYYO
20	OOOOOO	GWOOOO		WOOYWY	OWYOWO
21				YOOOOO	OWWOOO
22				OWWOOW	WWOOWO
23				WYYYYY	OYYWWY
24				WWOOYO	
25				WOOOOO	
26				WOYOOO	
27				WOOYOY	
28				WOWOYO	
29				WWWOOO	
30				YOWOOO	