The Loss of Intuition - A Lesson for the School Teacher?

Flavia R Jolliffe - Surrey, England

1. The study

Although there is a long tradition of research into concepts and intuitions regarding randomness and probability, few studies have been undertaken amongst United Kingdom university students. Thus the research done at Brunel University over the period 1984 to 1987 using a self-completion questionnaire to investigate the intuitive ideas concerning probability held by first year undergraduates was partly exploratory in nature. The 1984 study was a trial run and is not discussed in this paper. The questionnaires used in 1985, 1986, and 1987 were fairly similar to one another and took about fifteen minutes to complete during statistics tutorials taught by the author in the first few weeks of the university session.

The questionnaires started with an introductory note explaining what was expected, and ended by asking the respondent's sex and if probability had been studied before the start of the current universty session "not at all", "a little", or "in considerable depth". The questions on probablity were modelled closely on questions in Freedman et al. (1978), in Shaughnessy (1983), and on some used in studies reported in Green (1983, 1988). There were also variations on questions asked in the 1984 study on whether or not men have more sisters than women on average (Falk, 1982), and on the chance that the other child in a two-child family is a boy/girl. These will be discussed in a separate paper. The questions were presented in multiple choice format and many were followed by a question asking why the response given had been made. Care was taken to write in everyday language which meant that it was not possible to state the assumptions usually made in probability problems, such as independence of events.

Some details of the groups in the study are shown in Table 1. Most of the students were aged eighteen or nineteen years and 29 of the 136 students in the three years 1985 to 1987 were female. Very few students on the Computer Systems degree course had studied mathematics beyond the age of sixteen and students studying for degrees in Chemistry or Materials Science were weak in mathematics. All other students were specialising in some mathematical subjects and had a strong foundation in mathematics.

TABLE 1
Some summary details of the groups in the study

Year	Degree Course	No. students in study	No. who had studied probability before start of session
1985	Computer Systems	26	21
	Maths and Management	13	13
1986	Computer Systems	25	13
	Maths *or* Stats	23	18
1987	Computer Systems	36	20
	Chemistry *or* Materials Science	13	8

More than two-thirds of the students in the three years combined said that they had studied some probability before the start of the current session (see Table 1). Hardly any of these said that their study was in considerable depth. In addition, in 1986 students in mathematics or statistics degree courses had had several lectures on probability at university before taking part in the study. This should be borne in mind when considering the results, as it is likely that a student who has studied probability will be aware that there are correct answers to questions involving chance and will in consequence respond to a questionnaire in a different manner to probabilistically naive students. The fact that some students appeared to be taking the questionnaire rather more seriously than was warranted and that some rushed as if there were a time constraint might also have affected responses.

2. Some results

Classification of responses to the probability parts of questions was relatively straightforward as the number of different responses was either set by the multiple choice format or small. In contrast, classification of responses to the why parts of questions involved some ingenuity as responses did not always fit neatly into categories and were sometimes worded in such a way that considerable interpretation as to what was meant was needed. In a few cases they could not be classified at all. Possibly students were not sure how they were meant to respond. Responses of "don't know", "guessed", and blank responses were grouped together as all were thought to indicate that the respondent did not know, or could not formulate, the reason why a particular response had been given. On the probability parts of questions students tended to answer in the same way as one another, but there was less uniformity as regards the reasons for that response, and by no means all modal responses were correct.

Questions where nearly all students give the correct response and most give the correct reason for that response sometimes give little insight into intuition. For instance, in 1985 the first question asked which number was hardest to throw with an ordinary 6-sided die or whether they were all the same, and 37 out of 39 students thought all numbers were the same. One of the other two students said 6 was hardest and the other that 6 and 1 seem harder. Thirty-one students justified their answer either by referring to the physical property of the die or by using an equal probability type of argument. Possibly all that can be concluded is that a question of this type is not suitable for testing probabilistic intuitions with this age group. Even if they *think* 6 is harder to throw they have *learnt* that it is the wrong response.

Similarly on a question requiring a comparison of the chance of drawing a red marble from a box containing 1 blue and 9 red marbles with the chance of drawing a red marble from a box containing 10 blue and 90 red marbles, almost everyone responded correctly and could correctly state why. One possible risk of asking such questions is that students might think them trivial and so not pay full attention to later, more searching, questions. Both the die question and the marbles question had been omitted by the time of the 1987 study.

Three questions on tossing coins were used in each of 1985 to 1987, with some minor variations each year in an attempt to improve presentation. The coin was described as an ordinary coin in 1985. In 1986 this was changed to a fair coin in the hope that this would make clearer that heads and tails were equally likely. However, one student asked whether this meant that the coin was used in a fairground, so in 1987 the description was changed to unweighted. All three terms ordinary, fair, and unweighted, are used in texts and can be interpreted in terms of probabilities, but they cannot necessarily be assumed to be understood in this way by those with no formal knowledge of probability. Students might wonder about the "real" meaning of the description in a way that younger children would not, and students who have had some instruction in probability might expect an indication of whether the probability of a head can be taken to be 0.5 so that just referring to "a coin" would not be satisfactory. A layman's description of independent events is even more tricky. How to set up the experimental conditions in a manner appropriate to this age group when there is no common background in probability is a problem which occurs in many of the questions customarily asked in studies of this kind and one which has not been given much attention.

On a question concerning a coin which had been tossed five times with heads appearing every time almost everyone thought that a head was as likely as a tail on the next toss and could give an acceptable reason why, with only a few saying a tail was more likely on the next toss because it was time for a change. However, not all students who demonstrated an apparent understanding of coin tossing here carried this through to the next question comparing the probabilities of the two ordered sequences HTTHTH and HHHHTH. Approximately 30% of students did not give the correct response here, nearly all of these saying that the order HTTHTH is more likely than the order HHHHTH, with one or two saying the reverse and a few "don't know". The reasons for responses were sought in 1987 only, but the numbers are too small to make any conclusions.

If the responses to the first two questions on coins suggested that the groups as a whole had a reasonable understanding of coin tossing, the responses to the third question almost suggested the opposite. This question asked if 2 or 100 tosses of a coin gives a better chance of a win where a win occurs if the number of heads is exactly equal to the number of tails. Some results are shown in Table 2. More than half the students gave an incorrect response (three-quarters of the students in 1986) and those who had studied some probability did no better than the others. Reasons for the responses were obtained in 1987, but were not very enlightening, consisting mainly of "don't know", statements that P(H) = P(T), or reasons where it was hard to interpret the English, let alone the thought process. It has to be said that students at Brunel are not strong in written English, but it can also be conjectured that some students who have studied probability find the language of probability incomprehensible and some of them might have been responding in what they thought were correct terms.

TABLE 2

A - coin will be tossed either 2 times or 100 times.

You will win £5 if the number of heads is exactly equal to the number of tails.

	1985		1986		1987	
	a	b	a	b	a	b
2 tosses gives better chance of a win	16	11	12	10	18	13
100 tosses gives better chance of a win	9	9	15	9	14	5
Chance of winning same on 2 as 100 tosses	13	13	20	16	16	9
Don't know	1	1	1	1	1	1

a Total number making response.
b Number making this response who had studied probability (includes study at university for 1986 group).

The questions on coins were followed by two questions involving the sex of newborn babies, both questions having some similarity to the question asking for a comparison of the magnitudes of the probabilities of one head in two tosses of a coin and fifty heads in a hundred tosses. In all three questions consideration of the surprise associated with each event leads to the correct response. As has been seen, this last question on coins was not well done, and when it comes to the questions on babies the picture worsens. Over the three years only six students out of one hundred and thirty-six answered the two questions on babies and the similar question on coins correctly.

In 1985 the questions on the sex of babies had no run in, in 1986 they each started with a statement *The chance of having a baby boy is about 1/2*, and in 1987 *About 50% of all babies born are boys.* Looking at the 1985 results it was clear that some information about the probabilities involved was needed, especially for those students who had had some drilling in answering probability questions. However, the statement used in 1986 was in some ways too technical, and the imprecise way in which the probability was stated might have worried some students. In addition it avoided the issues of whether the chance is the same for all couples and whether there is independence between sexes within the family, on both of which points students have strong views so that they find it hard to step back and take an overall view of the sex distribution. This is possibly something which does not happen with younger respondents. *About 50% of all babies born are boys* contains sufficient information for the situations under consideration without drawing attention to finer details which are not important here.

The first question on babies asked which was more likely - 7 or more boys out of the first 10 babies born in a new hospital or 70 or more boys out of the first 100 babies, or whether they were equally likely. Following Green, the option "No-one can say" was also offered. This question should not have presented any difficulties as regards understanding the problem set-up. The pattern of responses in all three years was similar, although there are some important differences. The modal response in each year was that "7 or more boys out of 10" and "70 or more out of 100" were equally likely, many students justifying this by saying the proportions were the same. In 1986 half the responses were "same", 17 of these coming from the 23 students who had had several hours' probability study at university prior to completing the questionnaire. The proportion of correct responses increased each year from 5 out of 39 in 1985 to 12 out of 46 in 1986 to 15 out of 49 in 1987. Part of this increase could be due to the changes made in the way information was given about the probability that a boy is born. Choice of "no-one can say" by 16 of the 39 students in 1985 (compared with 17 choosing "same") can be partly explained by the lack of information that year about the probability of a boy, although not all students gave this as a reason for that response. In 1986 and 1987 a quarter of the students elected for "no-one can say". It was very noticeable in all three years that students choosing this response gave such reasons as "only nature can tell" and "there are many factors which determine babies". The numbers thinking that 70 or more boys in 100 was more likely than 7 or more boys in 10 were negligible in each year.

The second question on babies was based on one used by Shaughnessy (1983). In 1985 the wording used was: *Over an entire year, would there be more days when at least 70% of babies born were boys in ...* and the options given were *a large hospital, a small hospital,* and *the size of the hospital makes no difference.* In 1986 the question started with some scene setting in the hope that students would then be forced to think about individual days. Thus: *The chance of having a baby boy is about 1/2. In two hospitals, one large, one small, a daily record is kept of details of births. One of the summary measures found is the number of boy babies born in the day as a percentage of the total births during the day.* It then continued as in 1985, but referring to the large/small hospital in the response options. One of the problems about scene setting, which perhaps occurs here, is that it sometimes introduces unnecessary complications. Although the summary measure described is relevant, in thinking through the meaning of the sentence defining it students might lose sight of what the question is asking. In 1987, then, that sentence was deleted, and the first sentence was *About 50% of all babies born are boys.*

In each year the majority of students thought the number of days would be the same in a large hospital as a small, as many as 31 out of 39 choosing this in 1985. In many cases the reason given was that as the question was concerned with a percentage the size of the hospital did not matter, which suggests students were thinking about the proportion of boys born over an entire year and ignoring the reference to days. Another popular reason given was that boys and girls were equally likely, and other reasons were that the size of the hospital did not affect the probability of a boy and that a mother did not choose the hospital according to the sex of her baby. Several students could not give any reason for their response. As in the previous question, the proportion choosing the correct response increased each year - from 3 out of 39 in 1985 to 9 out of 48 in 1986 to 16 out of 49 in 1987 (but not all those who responded correctly on one question responded correctly on the other also) while the proportion choosing "same" decreased. Only two or three students each year chose the option "large hospital".

These results are slightly different to Shaughnessy's (1983). This might be because of the difference in wording used or perhaps because of a culture difference as regards sizes of hospitals in the USA and in England. Evans and Dusoir (1977) report an increase in the proportion of correct answers when the form of this question was simplified. The position of a question in a questionnaire might also make a difference. Although there is little evidence of it in the results discussed here, there might be a learning effect when several similar questions are presented together. Students did not appear to realise that coin tossing could be used as a model for the sexes of babies, or perhaps they felt it was inappropriate. It would be nice to think that the improved results each year were due to changes in wording. A similar improvement was not shown on the similar question on coins (Table 2).

3. Conclusions

The research reported here into the intuitive ideas about probability held by first year undergraduates studying mathematics or other scientific subjects suggests that experience, maturity, and to some extent instruction in probability, have a greater influence on the way in which students respond than does intuition. This will not necessarily be true in other subgroups of adults. It might be more appropriate to investigate what students already know about proportions and chance than their intuitions, and the joint study with Fay Sharples of the University of Waikato, New Zealand (described in a Miscellaneous session at ICOTS 3) has this as one of its aims. In looking for intuitions we might be looking for something that is not there. Obviously the way in which questions are presented and the exact form of the wording affects responses, but there was some evidence in the studies that question wording which is suitable for use with children may be less satisfactory with older age groups.

The tendency for those students who had had more instruction in probability to give wrong responses more often than students who had had little or no instruction, which was particularly noticeable in the 1986 students who had been given several lectures on probability (but not from the author) just before taking part in the study, is disturbing. Were students applying their new knowledge of probability wrongly? Were they so confused by the whole topic that they saw complications in the questions which were not there and perhaps thought that when intuition led to a correct response it must be wrong? Or was it simply, as others have found, that previously held misconceptions are unaffected by instruction? Without further research such as in-depth interviews one can only guess, but what is certain is that probability teaching at universities in the United Kingdom has traditionally been mainly theoretical. Introduction of practical experience and simulation of probabilistic situations seems desirable and could well give students a better understanding and feeling of the topic (Goodall and Jolliffe, 1988).

References

Evans, J St B T and Dusoir, A E (1977) Proportionality and sample size as factors in intuitive statistical judgement. *Acta Psychologica* **41**, 129-137.

Falk, R (1982) Do men have more sisters than women? *Teaching Statistics* **4**, 60-61.

Freedman, D, Pisani, R and Purves, R (1978) *Instructor's Manual for Statistics*. W W Norton and Company, New York and London.

Goodall, G W and Jolliffe, F R (1988) The training of Brunel University undergraduates who intend to become statistics teachers. In: R Davidson and J Swift (eds) *Proceedings of the Second International Conference on Teaching Statistics*. University of Victoria, Canada, 137-140.

Green, D R (1983) A survey of probability concepts in 3000 pupils aged 11-16 years. In: D R Grey, P Holmes, V Barnett, and G M Constable (eds) *Proceedings of the First International Conference on Teaching Statistics, Vol II*. Teaching Statistics Trust, University of Sheffield, England, 766-783.

Green, D R (1988) Children's understanding of randomness: Report of a survey of 1600 children aged 7-11 years. In: R Davidson and J Swift (eds) *Proceedings of the Second International Conference on Teaching Statistics*. University of Victoria, Canada, 287-291.

Shaughnessy, J M (1983) Misconceptions of probability, systematic and otherwise; teaching probability and statistics so as to overcome some misconceptions. In: D R Grey, P Holmes, V Barnett and G M Constable (eds) *Proceedings of the First International Conference on Teaching Statistics, Vol II*. Teaching Statistics Trust, University of Sheffield, England, 784-801.