Reasoning and Expressing Probability in Students' Judgements of Coin Tossing

Jonathan Moritz
University of Tasmania
<Jonathan.Moritz@utas.edu.au>

Jane Watson
University of Tasmania
<Jane.Watson@utas.edu.au>

A survey item based on a newspaper article about coin tosses was administered to 1256 students from grades 6 to 11. Few students determined the probability of four successive tails. Most students considered that heads and tails were equally likely for a subsequent fifth toss, often describing the probability as "50-50". Students in higher grades were more likely to respond appropriately. Results are discussed with reference to equiprobability, independence, the gambler's fallacy, and the outcome approach to probability.

> Australian cricketers lost in darkest Africa ... Border wanted to bat, but he has now lost eight of nine tosses in international matches on tour ... When he retires soon, his new sporting life will not, presumably, be as a two-up player. ("Lights, cam and a fair bit of inaction", *The Mercury*, 6 April, 1994, p. 52)

Newspapers extracts such as this illustrate that understanding probability is an advantage in various social contexts including sport, gambling, and the daily news. In these contexts, events such as a sequence of losing coin tosses often provoke intuitive comments of "that's not fair" or "how unlucky". "The language of chance is widely used in a colloquial way... students should be helped to refine and extend their use of this language so that they are more able to make sense of their everyday experiences" (Australian Education Council [AEC], 1991, p. 163). Possible activities for upper primary school students (Band B) include "analyse simple experiments (e.g. those involving single coins, dice and simple spinners, and equally likely events); make a systematic list of possible outcomes and assign simple numerical probabilities based on reasoning about symmetry" (AEC, 1991, p. 170). Also related to coins, for higher grade levels, the concept of independence appears in the curriculum for secondary students (Level 7), evident when a student "estimates probabilities... and assigns probabilities using complementarity and independence", such as to "calculate the probability of getting at least one head in five tosses of a coin using 1 – [minus] 'the probability of getting no heads in five tosses'" (AEC, 1994, p. 124).

The current study investigated school students' probability judgements concerning coin tossing sequences. Of particular interest were (a) the expressions of probability that students used indicating whether they used intuitive estimates or calculations to determine probabilities, (b) the incidence of responses indicating reasoning about equally likely outcomes and consistency with the concept of independence for coin tossing.

Reasoning and Expressing Probability

Much research of probabilistic reasoning has concerned errors people make based on intuitive misconceptions or heuristics (Fischbein & Schnarch, 1997). The *representative heuristic* is a belief that even small samples of data should reflect the parent population (Kahneman & Tversky, 1972). This heuristic explains the *gambler's fallacy*, a negative recency expectation; for example after a sequence of heads in coin tosses, tails is expected to even up the numbers of heads and tails. Students may exhibit inconsistencies in reasoning as they switch between reasoning based on intuitions or based on mathematical concepts (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993).

Truran and Truran (1997) suggest that students' responses to many probability questions should be interpreted with reference to the particular random generator involved and to the wording of the question, as students' intuitions and reasoning depend on their interpretation of the question wording. Typical questions focus on which outcome "is more likely", "will occur more often", or "is easier to get". Some students, however, appear to interpret most probability questions as what "you think will happen", that is prediction of an individual outcome (Konold, 1989; Konold et al., 1993). Known as the *outcome approach*, students interpreting probability questions in this way take one of three stances in their prediction: the outcome will happen, the outcome will not happen, or the outcome cannot be predicted. Metz (1997) further observed that some students believe that chance situations involve patterns-without-uncertainty (over confidence in predicting yes or no), and others believe in uncertainty-without-patterns (no basis for prediction), whereas the educational goal is an appreciation of quantifiable patterns of uncertainty.

Students' expressions of probability may indicate how they interpret a question. Results from other studies of development of chance measurement for students of grades 3 to 11 (Watson, Collis, & Moritz, 1997; Watson & Moritz, 1998) have indicated that younger students often use words or non-normative numerical expressions for measuring chance. Even an expression that appears to indicate a numerical value, such as "50-50", may be a colloquial expression indicating basic uncertainty but not measurement of chance (Konold et al., 1993). Thus any research evidence concerning students' reasoning with regard to equally likely outcomes or independence should be considered in the light of the problem situation, the wording of questions, and the expressions students offer.

Equally Likely Outcomes and Independence

The belief in equally likely outcomes for coins and dice is not universal. Kerslake (1974) found that 68% of grade 1 students and 59% of grade 4 students believe some numbers on dice are easier to get. Green (1983a) also found that only 45% of 11-12 year olds and 74% of 15-16 year olds believe there is no difference in whether a red or green face of a counter is more likely when tossed. When asked whether a 1 or a 6 on a die was easier to throw, Watson et al. (1997) found that 55% of grade 3 students and 87% of grade 9 students chose "equally easy", but fewer could provide a simple justification.

On the other hand it is possible to exhibit an *equiprobability bias* (Batanero, Serrano, & Garfield, 1996): an over-generalisation of equally likely outcomes to inappropriate situations. Lidster, Watson, Collis, and Pereira-Mendoza (1996) found that school students tended to assume dice were fair even after gathering data that suggested a heavy bias. Green (1983b) asked students about the outcomes of 100 drawing pins landing up or down, and found that around 60% of secondary students at any grade level chose the option "all these results have the same chance" although other multiple choice options more clearly reflected the given data of a previous trial that 68 pins landed point up and 32 landed point down.

Turning to independence, events A and B are *independent* if and only if $P(A|B) = P(A)$; that is, if one assumes tosses of a coin are independent, one expects that the likelihood of getting heads or tails is not affected by knowledge of previous tosses of the coin. Results of research studies, summarised in Table 1, have indicated that most secondary school students consider heads and tails equally likely for the next toss of a normal coin following a sequence of successive outcomes, though a significant number of students indicate a preference for a negative recency outcome in accord with the gambler's fallacy. Konold et al. (1993) described apparent inconsistencies in student reasoning, and distinguished between reasoning from representativeness that a *sequence* of five outcomes should display a balance of heads and tails, and reasoning about *the fifth outcome* that has two equally likely outcomes. As Truran and Truran (1997) point out, "given that most naive students do not have the combinatorial skills to assess these probabilities [of several five-toss sequences], such a response [based on representativeness] is not surprising" (pp. 94-95).

Table 1
Results of Four Studies of Students' Preferred Outcome in Coin Tossing Situations

Research Study	Previous event	Student response (H, T, or =), Grade/Age, and Proportion of response
Green (1982)[a]	5 Heads	(=) 11-12yo, 67%; 15-16yo, 80%
Brown, Carpenter, Kouba, Lindquist, Silver, and Swafford (1988) [b]	4 Tails	(=) grade 7, 47%; grade 11, 56% (H) grade 7, 38%; grade 11, 33%
Konold (1989), Konold et al. (1993)	4 Heads	(=) high school, 79%; uni graduates, 96%
Fischbein and Schnarch (1997)	3 Heads	(=) grade 5, 7, 9, 11: 40%, 55%, 70%, 90%
		(T) grade 5, 7, 9, 11: 35%, 35%, 20%, 10%

[a] *Note.* [a]Cited by Borovcnik & Bentz, 1991; [b]cited by Konold et al., 1993.

Research Questions

The current study of school students' responses to coin tossing outcomes investigated students' estimates and expressions of probabilities. When asked about a sequence of four tails, would students (a) calculate the probability of the sequence, (b) estimate the chance as less than 0.5 to indicate at least an intuitive sense for the measurement of chance of the sequence, or (c) use an outcome approach of basic uncertainty such as "50/50"? When asked about a subsequent toss, would students consider heads and tails equally likely, consistent with the concept of independence, and would they assign probabilities of 0.5 for each outcome? It was of interest whether appropriate responses would be more likely from students in higher grades or from those students who appropriately measured chance on traditional probability items concerning simple events.

Method

Survey responses (N = 1581) were gathered from 1256 students in grades 6 to 11 at 20 government primary schools, secondary schools, and secondary colleges distributed around Tasmania. Curricula in these schools were individually adapted in general adherence to the Australia-wide curriculum (AEC, 1991). During 1995, 206 grade 6 students, 301 grade 8 students, 294 grade 9 students, and 72 grade 11 students were surveyed; during 1997, 80, 206, and 39 of these students, then in grades 8, 10, and 11 respectively, were surveyed again. Responses were also obtained from additional students during 1997: 166 in grade 6, 74 in grade 8, 90 in grade 9, 46 in grade 10, and 7 in grade 11. Approximately equal numbers of females and males were surveyed at each grade level in each year.

A survey item, shown in Figure 1, was developed involving four parts closely related to the interview protocol of Konold et al. (1993). This item was last on a 10-item survey involving newspaper extracts about chance and data concepts (Watson, 1994). The survey was administered to whole class groups during 45 minutes of class time. Students who did not respond to the item, due to time or inclination, were excluded from analysis.

> During the recent Australian cricket tour of South Africa, the Hobart Mercury (6/4/1994, p. 52) reported that Allan Border had lost 8 out of 9 tosses in his previous 9 matches as captain. Imagine his situation at this point in time.
>
> *(a)* Suppose Border decides to choose heads from now on. For the next 4 tosses of the coin, what is the chance of the coin coming up tails (and him losing the tosses) 4 times out of 4?
>
> *(b1)* Suppose tails came up 4 times out of 4. For the 5th toss, should Border choose
>
> ❑ Heads ❑ Tails ❑ Doesn't matter
>
> *(b2)* What is the probability of getting heads on this next toss?
>
> *(b3)* What is the probability of getting tails on this next toss?

Figure 1. Survey item (italics denote labels not shown in original survey item).

Responses to part (a) were coded into four categories: value < 0.5, value = 0.5, other value (value > 0.5 or indeterminate), and no response. Common response expressions were also identified, in particular distinguishing various expressions of the probability value 0.5. Responses to part (b1) were coded as "H", "T", or "=" according to multiple choice selections "Heads", "Tails", and "Doesn't Matter" respectively; in a few cases where no choice was made, the choice was inferred from parts (b2) and (b3). Responses to parts (b2) and (b3) were coded either as both values = 0.5, or as other values (including "chance" or "luck", indeterminate values, differing values, or the same values but $\neq 0.5$). For part (a) and for part (b)[1], analysis was conducted to examine if there were any significant differences between students in comparable grades who were (i) repeating and non-repeating students (grades 8, 10, and 11), (ii) females and males (grades 6, 8, 9, 10, and 11), and (iii) drawn from 1995 and 1997 samples (grades 6, 8, 9, and 11). Of 24 tests in total, only two were significant with $p <$ 0.01: for grade 10 responses to part (a), appropriate responses with a value less than 0.5 were more common from males ($\chi^2_3 = 12.93$, $p = 0.0048$) and from non-repeating students ($\chi^2_3 = 12.93$, $p = 0.0048$). As these differences were of marginal significance for the number of tests, response data were combined across repeating and non-repeating students, females and males, and samples from 1995 and 1997.

Results

In response to part (a), few students at any grade level appropriately evaluated the chance of four successive tails as 1/16, as shown in Table 2. Many students, particularly from grade 9 or below, did not respond. Of those who did respond, most responded with a value of 0.5. Those responding with a value of 0.5 often expressed variants of "50%", "50-50", "50/50", "2/4", "2 out of 4", and "2 times out of 4". The first three expressions are consistent with a valuation of 0.5, but are also consistent with the outcome approach that "it is possible"; for example, a grade 8 male responded, "it's both half & half because you can't predict what's going to come up". The expression "2/4" may be a description of expecting 2 tails out of 4

[1] In χ^2 tests for part (b), "H" and "T" responses were combined to form four outcome categories: "=, each value = 0.5"; "=, other values"; "H/T, each value = 0.5"; and "H/T, other values".

tosses, or from various thinking about 1/2; for example, one grade 11 female responded: "P(half + half + half + half). The chances of Border losing the tosses 4 times out of 4 are 4/8 or 2/4. There is a 50-50 chance he will lose on each of the 4 tosses." Students in higher grades were more likely to respond with values less than 0.5; common expressions included "25%", "1/4", "12.5%", "1/8", and various chance words such as "unlikely". One grade 11 female responded "1/2 + 1/2 + 1/2 + 1/2 = 1/8 chance of tails four times out of four", and another responded, "1 in 5 chance {HHHH TTTT HHHT TTTH TTHH}". Some students gave values greater than 0.5 such as "3/4", "4/4", and "probable". Many responses were of indeterminate value, such as chance expressions that did not clearly indicate equality or inequality with 0.5, for example "anything is possible" and "doesn't matter, it's the luck of the toss". Other expressions included "no", "yes", "2/2" (unknown intended probability), and "there is equal chance of either coming up" (the student may have been considering a single toss rather than a sequence of four tails).

Table 2
Percentage Responses to Part (a) by Grade (N = 1581)

Response category	Grade				
(Response expression)	6	8	9	10	11
Value < 0.5 (1/16[a])	0	0	3	5	16
Value < 0.5 (Other # or word)	12	9	11	12	13
Value = 0.5 (50%)	8	16	12	17	19
Value = 0.5 (50-50)	7	11	11	23	7
Value = 0.5 (2/4)	11	7	7	6	11
Value = 0.5 (Other # or word)	2	1	3	3	5
Value > 0.5 or Indeterminate	19	13	15	13	14
No response	41	42	38	21	15
n	372	455	384	252	118

Note. [a]Includes two responses of "1/2 x 1/2 x 1/2 x 1/2" where this expression was incorrectly evaluated.

In response to part (b1), most students selected "=", as shown in Table 3. Students in higher grades were less likely to respond "H" or "T", and more likely to respond "=" with values of 0.5 for parts (b2) and (b3). Those responding with values of 0.5 often expressed variants of "50%", "50-50", "50/50", or "1/2". Many students gave inconsistent responses, combining "H" or "T" in part (b1) with values of 0.5 in parts (b2) and (b3). Many responses of indeterminate value were due to non-response to parts (b2) or (b3).

Responses to part (b) differed according to responses to part (a). Of students' responses to part (a) with (i) a value less than 0.5 (n = 214), (ii) a value of 0.5 (n = 569), and (iii) other responses or no response (n = 798), the proportions of students responding "=, each value 0.5" to part (b) were 59%, 72%, and 38% respectively. Hence 8% (n = 1581) of the sample responded in the correct response category on both parts.

Table 3
Percentage Responses to Part (b) by Grade (N = 1581)

Response to part (b1), values for parts (b2) and (b3)	Grade				
	6	8	9	10	11
=, Each value = 0.5	32	55	52	73	79
=, Other values [a]	30	20	25	14	9
H, Each value = 0.5	5	6	5	3	3
H, Other values [a]	16	10	8	6	2
T, Each value = 0.5	6	4	5	3	3
T, Other values [a]	12	6	6	2	3
n	372	455	384	252	118

Note.[a]Responses include "chance" or "luck", indeterminate values, differing values, or the same values but ≠ 0.5.

A subset of responses (n = 1497) were matched to the students' chance measurement levels (Watson & Moritz, 1998). Levels were based on responses to three items: (1) chance of outcomes when a die is rolled, (2) chance of outcomes drawn from a bag, and (3) comparison of chances of drawing one colour of marble from two boxes with the same ratio of colours. Level 0 indicated responses not related to judgements of uncertainty. Levels 1, 2, and 3 indicated basic uncertainty, a qualitative judgement, and a quantitative judgement, respectively. Levels 4 to 6 indicated proportional reasoning in the third item, with higher levels for more explicit numerical justification. The distribution of responses to parts (a) and (b) of the coin tossing item differed according to chance measurement level, as shown in Table 4 (part (a), $\chi^2_{21} = 145.3$, $p < 0.0001$; part (b), $\chi^2_{21} = 173.7$, $p < 0.0001$). Higher levels were associated with more appropriate responses to parts (a) and (b).

Table 4
Percentage Responses by Chance Measurement Level (N = 1497)

Response category for part (a)	Chance Measurement Developmental Level							
	0	1	2	2.5	3	4	5	6
Value < 0.5	20	8	8	8	11	17	13	35
Value = 0.5	20	8	29	34	41	45	42	34
Value > 0.5 or Indeterminate	20	23	23	16	15	13	15	6
No response	40	61	41	42	33	25	30	24
Response to part (b1), values for parts (b2) and (b3)								
=, Each value = 0.5	10	24	33	44	59	68	58	79
=, Other values [a]	30	40	33	25	19	16	26	9
H/T, Each value = 0.5	0	8	10	12	10	7	4	5
H/T, Other values [a]	60	27	23	20	12	8	11	6
n	10	62	177	358	270	436	53	131

Note.[a]Responses include "chance" or "luck", indeterminate values, differing values, or the same values but ≠ 0.5.

Discussion

For the event of four successive tails, less than 30% of students, even by grade 11, responded that with a probability less than 0.5, and very few could calculate the probability to be 1/16. Many students in grades 6, 8, and 9 gave no response, and many students in all grades expressed a value of 0.5 such as "50%", "50-50", or "2/4". These results indicate not only that students are poor at *calculating* the probability of compound events by multiplying probabilities of independent events, but also that they are poor at *estimating* such a compound probability using intuition. Many students appeared not to be used to probabilistic reasoning about compound events, and they based responses on the outcome approach (Konold et al., 1993) in that they were unwilling to predict the outcome. It is not clear, however, whether these students considered the outcome to be four tosses or one. It had been hoped that four tails would be a long enough run length to prompt students' intuitions to acknowledge that the sequence was not only possible but also *rare*. Perhaps for many students a more strikingly long run length is required to trigger this thinking. The dilemma associated with run length and equiprobability is illustrated in the following dialogue from an interview with a grade 9 male.

I: What is the chance of getting tails four times in a row?
S: Pretty good chance, oh not really, not really a pretty good chance. Just the luck of the flip.
I: Yeah ok. How likely, do you think would it be to get four tails in a row, what sort of number?
S: Oh, 50 percent I reckon... They both have the same amount of chance of coming out. [...]
I: What about 100 tails in a row, do you think that is very likely? [...] Would that still be 50 percent or would that be a different number, for probability of getting 100 tails in a row?
S: Oh... Not... One percent I reckon.
I: Right, it might be one percent. Do you think for four tails in a row, it is 50 percent, or...?
S: I suppose it would have a good chance of coming up.
I: Yeah. Would that mean it's 50 percent or a different number?
S: 50 percent I still reckon.

Concerning the fifth coin toss following four successive heads, the high incidence of correct multiple choice responses closely mirrored that found in small samples by Konold (1989), though the trend of improvement with increasing grade was less pronounced than that found by Fischbein and Schnarch (1997). Two aspects would appear to demand attention in the classroom: expressing measurement of chance, and acknowledging independence to reduce the gambler's fallacy. These two aspects may be related. For example, another grade 9 boy in an interview, responded to part (b1), "I reckon he'd have a better chance of doing heads, cause I reckon the next one would come up heads", then suggested 50% for parts (b2) and (b3). When the interview commented about heads being more likely, and thus whether the chance would be 50%, the student responded as follows.

S: I'm not sure... it would be fifty percent because you wouldn't know which is going to come up.
I: So do you think it really doesn't matter which one he picks, or do you think he is generally better off picking heads?
S: Oh it doesn't really matter, I suppose. I'll rub that one out... It's just the luck of the flip.

It is not clear whether the student's rejection of the gambler's fallacy is in favour of the outcome approach, that one cannot predict, or in favour of measuring the chance as 50%. A similar indeterminacy occurred in interpreting responses of about 10% of students, who were inconsistent by combining a response of "H" or "T" with values of 0.5 (see Table 3). Evidence of improved performance in relation to chance measurement level (see Table 4) gives some hope that understanding of traditional probability problems may assist students to reject both the gambler's fallacy and the outcome approach, in favour of measuring chance. Konold (1989) concluded by advocating use of asymmetric distributions to encourage chance measurement, along with the assertion that some outcomes are more likely than others, rather than symmetric distributions, which are consistent with the view that there is no "basis for prediction" as outcomes are equally likely. Introducing asymmetric binomial distributions, such as those resulting from dropping drawing pins and counting the two outcomes (point up or point down) (Green, 1983b), may be an important step to prompt students to measure chance numerically, whether in situations where empirical data may be required to provide chance measurement estimates, or in classical probability situations of equally likely outcomes. Further encouragement to measure chance numerically should reduce reliance on the outcome approach and on intuitions such as the gamblers' fallacy, and overall, tasks based on compound events should encourage students to combine reasoning about simple events with the concept of independence.

References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Vic.: Author.

Australian Education Council. (1994). *Mathematics - A curriculum profile for Australian schools*. Carlton, Vic.: Curriculum Corporation.

Batanero, C., Serrano, L., & Garfield, J. B. (1996). Heuristics and biases in secondary school students' reasoning about probability. In L. Puig, & A. Gutiérrez (Eds.), *Proceedings of the 20th Annual Conference for Psychology of Mathematics Education*, (*Vol. 2*, pp. 43-50). Valencia, Spain: PME.

Borovcnik, M., & Bentz, H. J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73-105). Dordrecht: Kluwer.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, 96-105.

Green, D. R. (1983a). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (*Vol. 2*, pp. 766-783). Sheffield, U.K.: Teaching Statistics Trust.

Green, D. R. (1983b). From thumbtacks to inference. *School Science and Mathematics, 83*, 541-551.

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

Kerslake, D. (1974). Some children's views on probability. *Mathematics in School, 3*(4), 22.

Konold, C. (1989a). An outbreak of belief in independence? In C. Maher, G. Goldin, & B. Davis (Eds.), *Proceedings of the 11th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 203-209). Newark, NJ: Rutgers University Press.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392-414.

Lidster, S. T., Watson, J. M., Collis, K. F., & Pereira-Mendoza, L. (1996). The relationship of the concept of fair to the construction of probabilistic understanding. In P. C. Clarkson (Ed.), *Technology in mathematics education* (pp. 352-359). Melbourne: MERGA.

Lights, cam and a fair bit of inaction. (1994, April 6). *The (Hobart) Mercury*, p. 52.

Metz, K. E. (1997). Dimensions in the assessment of students' understanding and application of chance. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 223-238). The Netherlands: The International Statistical Institute.

Truran, J., & Truran, K. (1997). Statistical independence – One concept or two?: Implications for research and for classroom practice. In B. Phillips (Ed.), *Papers on statistical education presented at ICME-8* (pp. 87-100). Hawthorn, Vic: Swinburne University of Technology.

Watson, J. M. (1994). Instruments to assess statistical concepts in the school curriculum. In National Organizing Committee (Ed.), *Proceedings of the Fourth International Conference on Teaching Statistics.* (*Vol. 1*, pp. 73-80). Rabat, Morocco: National Institute of Statistics and Applied Economics.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9*, 60-82.

Watson, J. M., & Moritz, J. B. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10*(2), 103-127.