THE CONJUNCTION FALLACY AND LONGITUDINAL DEVELOPMENT OF CHANCE EXPRESSION

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

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

Two survey items asking for estimates of probability or frequency of everyday events (A), (B), and their conjunction (A and B), were completed by 2719 school students in grades 5 to 11. Cross-sectional and longitudinal analyses revealed chance expression improved with grade, but no change in incidence of conjunction errors. Gender differences favouring males occurred for some grades. Comparisons with responses to other probability items indicated incidence of conjunction errors is independent of development of basic chance measurement.

The conjunction fallacy arises in contexts where probabilities are considered for two events and their intersection (conjunction). According to classical probability theory, if sets A and B are defined, it is necessary that the conjunction set (A and B) is a subset of A and of B, thus P(A and B) is necessarily less than or equal to both P(A) and P(B). Previous research has found that people often violate this principle using reasoning based on a conjunction fallacy. Tversky and Kahneman (1983) and subsequent researchers used problems in social contexts, often variations of a problem that involved a character description of Linda, with respondents asked to judge the likelihood that Linda has various occupations (A), hobbies (B), or both (A and B). Tversky and Kahneman (1983) found that around 90% of university students violated the conjunction rule under certain conditions. They considered that some respondents may be averaging P(A) and P(B) to arrive at P(A and B), or thinking of the causal relations between A and B, thus rating (A and B) as more typical than (A) and (B) for the character description given. Their major finding, however, was that fewer conjunction errors (only 11%) occurred for simpler questions requesting frequency rather than probability estimates. They interpreted this as evidence that responses to probability items were often based on reasoning of typicality and intentional meaning of terms, using a more general *representative heuristic*, whereas responses to frequency items were based on extensional referents, which are countable.

Later researchers trialed variations of Linda-type tasks. Variations of response format included tasks to estimate frequencies such as "how many people out of 100..." (Fiedler, 1988), to estimate likelihood from 0 to 10 (Fisk & Pidgeon, 1997) or from 0 to 100 (Fantino, Kulik, Stolarz-Fantino & Wright, 1997), and to order alternatives (Morier & Borgida, 1984); estimating frequencies consistently yielded fewer conjunction errors. Researchers also employed variations of the components to be estimated, such as P(A), P(B), P(A and B), and P(A or B) (Morier & Borgida, 1984), as well as changing the social setting of the question to a traditional probability setting of marbles in urns (Yates & Carlson, 1986); these studies indicated that focusing attention on the components and the probability setting also reduced conjunction errors.

Only two of the research studies of the conjunction fallacy found involved school students. Fischbein and Schnarch (1997) asked students to rank P(A and B) and P(A) in a Linda type problem, and found 85% of grade 5 students committed the conjunction fallacy, whereas only 40% of grade 11 students did. Davidson (1995) also used Linda-type problems, asking young students to rate likelihoods on a 5 point scale. Conjunction errors were made by 35% of grade 2 students, rising to 57% of grade 6 students, apparently indicating increasing use of the representative heuristic with grade. No clear developmental trend emerges from these two studies, however the sharply contrasting results again draw attention to task variations of the form of expression for responses as important in accounting for incidence of conjunction errors.

Results from other studies of development of chance measurement for students of grades 3 to 11 (Watson, Collis & Moritz, 1997; Watson & Moritz, 1998) and odds for grades 6 to 11 (Moritz, Watson & Collis, 1996; Moritz, 1998) have indicated that younger students often use words or non-normative numerical expressions for measuring chance, and that sex differences favouring males are evident at selected grade levels in secondary school. Thus the expression of response to open-ended tasks and sex differences may contribute to the mixed results observed in the two studies of school students and shed light on the development of conjunction fallacy reasoning for school students.

The current study of school students' responses to conjunction problems investigated not only the incidence of conjunction errors, but also the ways that students express likelihood estimates. As the conjunction fallacy items were part of a larger survey, it was not possible to include the range of formats for questions used by all earlier researchers. Based on the experiences of other researchers it was decided to use an open-ended format asking for estimates. This format allows for a distinction to be made between responses that are appropriate, responses that explicitly express the conjunction fallacy, and responses that are undefined in terms of the context set. Rather than using in-school contexts such as marbles in urns, out-of-school contexts were used in an effort to explore student reasoning in everyday contexts and to reflect the application goals of the school curriculum. Three research questions were of interest. (1) In everyday contexts, do school students interpret two events and their conjunction in an appropriate fashion? What alternative interpretations arise and in what formats are responses offered? (2) Does student performance on these tasks improve with age, or differ between cohorts or sexes? (3) Is there an association between performance on conjunction estimation items and performance on other chance measurement items?

METHOD

Participants

Survey responses were gathered from 2719 students at 20 government primary schools, secondary schools, and matriculation colleges distributed throughout Tasmania. Responses were collected in 1993, 1995, and 1997, and totalled 3730 responses, including 785 responses from the same students surveyed again after a two-year interval, and a further 113 responses from students surveyed three times, in 1993, 1995, and 1997. The numbers of students surveyed from different schools varied across years and grades due to availability of students (see Table 1). Approximately equal numbers of males and females were surveyed in each year at each grade level. More details of the cross-sectional and longitudinal aspects of the larger study are found in Watson and Moritz (1998).

Items and Procedure

Two short answer items, shown in Figure 1, were adapted from those of Tversky and Kahneman (1983, p. 309). Both items required subjective likelihood estimates of everyday events, Item 1 in probability form and Item 2 in frequency form. For each item, it was expected that part (b) would be estimated as more likely than part (a). Item 1(c) involved the word "causing", restricting the likelihood of the event even moreso than the conjunction. These items were questions 15 and 18 of a 20-item chance and data written survey (Watson, 1994). The survey was administered to whole class groups during 45 minutes of class time. Some students who did not respond to these or later items, due to time or inclination, were excluded from this analysis.

Figure 1
Two Items Involving Conjunction Estimates

1. Please estimate:
(a) The probability that you will miss a whole week of school next year.
(b) The probability that you will get a cold next year.
(c) The probability that you will get a cold causing you to miss a whole week of school next year.
2. A health survey was conducted in a sample of 100 men in Australia of all ages and occupations. Please estimate:
(a) How many of the 100 men have had one or more heart attacks.
(b) How many of the 100 men are over 55 years old.
(c) How many of the 100 men both are over 55 years old and have had one or more heart attacks.

Coding and Analysis of Responses

Responses were entered into a spreadsheet in the form students wrote them. Responses to the two items were then coded according to (1) the numerical relation between the three numerical estimates of the probability or frequency for the three parts of each item and (2) the type of expression used. Numerical Relations included $c<min(a,b)$, $c=min(a,b)$, $min<c<max$, and $max(a,b)=c$, according to the numerical value expressed in part (c) in relation to the minimum and maximum numerical values expressed in parts (a) and (b). The distinction between $mix<c<max$ and $max(a,b)=c$ was made to further clarify the reasoning used by students, in particular to see if students might be using an averaging process (Tversky & Kahneman, 1983; Fantino, et al., 1997). The $c<min(a,b)$ and $c=min(a,b)$ categories were also coded as "correct", and $min<c<max$ and $max(a,b)=c$ as "fallacy". Word expressions were mapped to numerical values where possible: "unlikely" and "low" were assigned 0.25, "maybe", "medium" and "average" were assigned 0.50, and "likely" and "high" were assigned 0.75, with modifiers such as "very" being assigned more extreme values. An *undefined* numerical relation was entered for responses where a numerical value could not be reasonably inferred for all three parts to the item, or if word expressions to different parts could not be differentiated to determine which was intended to have a higher probability value. Categories of Expression included *frequency* (whole numbers 0-100), *percentage* (use of "%"), *fraction* (decimal, e.g., 0.25; part-whole ratio, e.g., 1/4; or part-part odds, e.g., 5/2), *yes/no* (simple use of either word), *word* expressing chance (e.g., "likely"), and *other* (e.g., "I don't know"). Expressions were also coded as Numerical (frequency, percentage, or fraction) and Non-numerical (chance word, yes/no, or other). If responses to all three parts used the same type of expression, this expression was assigned as the response expression for that item; otherwise *mixed* expression was assigned to the response.

Students' responses were analysed in three different ways. *(1) Cohort and cross-sectional analyses* using χ^2 tests involved the independent factors of cohort (1993, 1995, 1997), grade level, and sex. Responses from comparable grades collected from different cohorts were compared to investigate whether recent curriculum reform and implementation had affected incidence of the conjunction fallacy and students' expressions of chance. Responses of students from a cross-section of grades (and both sexes) were compared as one method of investigating conceptual development of students. *(2) Longitudinal analysis* was a second method for exploring conceptual development, in this case analysing differences in responses of 113 individual students gathered longitudinally over two 2-year intervals. *(3) Cross-item analyses* involved comparing students' responses to the probability (Item 1) and frequency (Item 2) forms. Responses were also compared to those of chance measurement tasks reported in previous studies (Watson, et al., 1997; Watson & Moritz, 1998) to explore understanding of other probability concepts that may impact on reasoning for conjunction tasks.

RESULTS

Cohort and Cross-Sectional Analyses

Tables 1 and 2 show the numerical relations and expressions used in responses to Items 1 and 2 respectively, for all 3730 responses grouped by grade and year of survey. The results reported include both those who were completing the items for the first time as well as those who were repeating the items in each year, as there were no significant differences in the numerical relations and expressions of non-repeating versus repeating students in comparable grades, with one exception. This exception was that repeating students in grade 7 were more likely to use numerical expressions in response to Item 1 (χ^2_2=17, p<0.001); this may be due to the additional non-repeating students at the grade 7 level being drawn from other feeder primary schools than the repeating students.

Table 1
Percentage Responses to Item 1 Coded by Numerical Relation and Expression

Response	1993 Grade		1995 Grade					1997 Grade						
category	6	9	5	6	8	9	11	5	6	7	8	9	10	11
Numerical Relation														
c<min(a,b)*	26	39	17	19	38	31	40	19	23	26	26	33	37	49
c=min(a,b)*	23	25	21	21	23	25	28	19	19	19	27	32	30	22
min<c<max^	11	15	5	12	13	16	19	11	13	15	20	13	15	14
max(a,b)=c^	7	7	6	8	7	4	7	6	6	5	7	8	7	8
Undefined	33	14	51	40	18	24	7	46	39	36	20	15	11	8
Expression														
Frequency	1	2	4	3	4	1	1	7	5	3	2	2	4	2
Percentage	23	38	15	29	45	36	49	14	23	30	44	44	45	37
Fraction	11	24	7	9	15	21	26	14	10	10	13	8	22	39
Word	19	14	18	13	11	14	7	15	22	16	17	26	12	6
Yes/No	16	5	24	16	6	10	1	20	18	12	5	6	2	0
Mix/Other	29	17	32	30	19	18	15	29	22	29	20	14	15	16
N	307	377	435	321	361	345	215	216	215	305	186	104	292	51

* denotes correct response categories, ^ denotes conjunction fallacy categories

The results for Item 1, shown in Table 1, were consistent for comparable grades across 1993, 1995, and 1997 cohorts. The only significant difference found was that grade 9 students in later years were less likely to use numerical expressions (69% in 1993 to 55% in 1997; χ^2_1=9.6, p<0.01). Combining the *c<min(a,b)* and *c=min(a,b)* categories in Table 1, the correct response rate improved over the grades, being fewer than half of students grade 7 or below, but more than half of students grade 8 or above. Many younger students used non-numerical expressions, resulting in the numerical relation between the three parts being undefined. Older students tended to respond with appropriate expressions of probability, such as percentage or fractional chance. The percentages of students making a conjunction error (combining *min<c<max* and *max(a,b)=c*) ranged from 11% to 27% across grade samples. The indication of increasing percentage of conjunction errors for older age groups, particularly evident in the grades surveyed in 1995, disappeared when undefined responses were removed from the analysis: the corresponding percentages of students making a conjunction error ranged from 22-34% across grade samples, with no trend across grades. More students responded in the *min<c<max* category than the *max(a,b)=c* category.

As for Item 1, the results for Item 2, shown in Table 2, were remarkably consistent for comparable grades across cohorts. The percentages of correct responses to Item 2 varied little across grades. Most students in all grades responded with expressions of frequency or of the percentage of the sample of 100 men. As a consequence, few responses had undefined numerical relations between parts. The results were also consistent across grades. The percentages of students making a conjunction error ranged from 29-44% across grade samples (32-51% when undefined responses were excluded). Of these, in contrast to Item 1, more students responded with *max(a,b)=c* than *min<c<max* .

Table 2
Percentage Responses to Item 2 Coded by Numerical Relation and Expression

Response	1993 Grade		1995 Grade					1997 Grade						
category	6	9	5	6	8	9	11	5	6	7	8	9	10	11
Numerical Relation														
c<min(a,b)*	38	46	39	36	42	43	40	38	44	33	31	41	38	51
c=min(a,b)*	16	18	14	18	16	16	17	18	14	10	20	17	19	14
min<c<max^	12	14	13	13	16	15	17	9	12	20	12	14	13	8
max(a,b)= c^	19	16	16	17	20	15	20	21	18	24	25	19	24	20
Undefined	14	6	17	15	6	11	7	14	13	13	12	8	5	8
Expression														
Frequency	54	58	51	45	51	53	47	56	52	46	34	48	52	57
Percentage	15	21	22	27	32	20	27	14	23	30	41	30	29	25
Fraction	6	7	2	4	5	6	7	7	4	5	6	6	8	6
Mix/Other	25	13	25	24	12	20	19	22	21	19	18	16	11	12
N	307	377	435	321	361	345	215	216	215	305	186	104	292	51

* denotes correct reponse categories, ^ denotes conjunction fallacy categories

Table 3 shows the performances of female and male students at each grade level (cohorts combined) in a consolidated form (*correct, fallacy, undefined*) for the two items. For Item 1, sex differences favouring males were observed in grade 9 ($\chi^2_2=11.6$, $p<0.01$) and grade 11 ($\chi^2_2=12.8$, $p<0.01$). For Item 2, there were no significant sex differences, despite the exceptionally high incidence of conjunction errors for grade 7 females.

Table 3
Percentage Responses to Items 1 and 2 by Grade and by Sex

Response	Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10		Grade 11	
category	f	m	f	m	f	m	f	m	f	m	f	m	f	m
Item 1														
Correct	38	37	44	44	39	49	54	63	55	67	65	68	59	78
Fallacy	10	16	20	18	24	16	26	19	24	17	26	18	33	15
Undefined	52	47	36	38	37	35	20	18	21	16	9	13	8	7
Item 2														
Correct	55	54	56	55	37	48	56	55	59	63	58	58	57	60
Fallacy	29	31	31	31	52	36	36	37	31	29	35	39	36	34
Undefined	16	14	14	14	11	15	8	8	9	7	8	3	8	6
N	317	334	432	411	150	155	273	274	425	401	146	146	143	123

Longitudinal Development

The results of longitudinal analysis of responses to Items 1 and 2 by 113 students in grades 6, 8 and 10 are shown in Table 4. Responses to Item 1 involved similar distributions to those shown in Table 1: many students in grade 6 used non-numerical expressions of the probability, and thus many have an undefined numerical relation, whereas older students more often used percentage expressions. There was an increase in correct responses after grade 6 but no evidence of change in rates of conjunction errors. Responses to Item 2 involved similar distributions to those shown in Table 2: most responses were expressed as frequencies or percentages, and there was a slight improvement in correct responses after grade 6. Similar rates of conjunction errors occurred across grades. To examine within-student longitudinal change, results were consolidated by grouping numerical relations into *correct*, *fallacy*, and *undefined*. In two 2-year intervals, 1993-1995 and 1995-1997 (the latter in parentheses for the following results) for Item 1, 60 (or 63) students remained in the same grouping, 14 (or 10) students improved their response from *fallacy* to *correct*, whereas 10 (or 15) reverted from *correct* to *fallacy*. Similarly, for Item 2, 53 (or 70) students remained in the same grouping, 21 (or 16) students improved their response from *fallacy* to *correct*, and 16 (or 14) reverted from *correct* to *fallacy*. Thus overall, the longitudinal study indicates that students developed in numerically expressing chance or frequency, but percentages of conjunction errors or correct conjunction reasoning were quite stable over time, with fluctuations of improvement and reversion in similar frequencies.

Table 4
Percentage Responses to Items 1 and 2 Assessed Longitudinally (N=113)

Response category	Item 1			Item 2		
	Grade 6	Grade 8	Grade 10	Grade 6	Grade 8	Grade 10
Numerical Relation						
c<min(a,b)*	26	45	39	33	48	46
c=min(a,b)*	25	21	27	21	13	23
min<c<max^	12	10	16	13	17	11
max(a,b)=c^	10	7	5	19	14	17
Undefined	28	17	12	14	8	4
Expression						
Frequency	1	4	5	58	48	61
Percentage	28	50	49	15	32	25
Fraction	13	11	17	3	5	4
Word	14	11	14	-	-	-
Yes/No	13	5	2	-	-	-
Mix/Other	30	20	13	22	14	10

* denotes correct response categories, ^ denotes conjunction fallacy categories

Cross-Item Analyses

Of 3730 responses, *correct* was similar for Item 1 (1946 responses) and Item 2 (2083 responses), *fallacy* was more common for Item 2 (1235 responses) than Item 1 (732 responses), whereas *undefined* was more common for Item 1 (1052 responses) than Item 1 (412 responses). Overall 31% of responses were *correct* to both items, and 8% were a *fallacy* to both. Of those who answered Item 2 either correctly or with a fallacy, 56% or 52% respectively answered Item 1 correctly. This difference, despite producing a significant χ^2 value, reflected little actual association in responses to the two items.

To explore the association between success on conjunction estimate items and more general understanding of chance measurement, 3616 responses to Items 1 and 2 were matched to developmental levels determined by Watson and Moritz (1998), scored from 0 to 6, based on responses to three chance measurement items earlier in the survey. These levels represent increasingly complex cognitive functioning evident across responses to three items concerning (1) likelihood of numbers occurring when a 6-sided die is rolled, (2) likelihood for an outcome drawn from a bag, and (3) comparisons of likelihoods of drawing one colour of marbles from boxes with similar ratios of colours. The distribution of responses to Items 1 and 2 by chance measurement developmental level is shown in Table 5. For Item 1, higher chance measurement developmental levels were associated with more *correct* and fewer *undefined* responses; *fallacy* rates were constant across developmental levels above level 2, beyond which at least simple quantification of chance measurement is evident. For Item 2, the *fallacy* rate was again independent of developmental level. The positive association with *correct* responses was less marked than for Item 1, corresponding to the weaker of association for this item of chance measurement level with increasing numerical expressions and with decreasing *undefined* responses.

Table 5
Percentage Responses to Items 1 and 2 by Chance Measurement Level

Response	Chance Measurement Developmental Level							
category	0	1	2	2.5	3	4	5	6
Numeric Relation (Item 1)								
Correct	22	33	41	46	57	64	62	68
Fallacy	10	13	17	20	22	21	20	22
Undefined	67	54	43	34	21	15	18	10
Expression (Item 1)								
Numerical	22	33	41	46	57	67	73	78
Numerical Relation (Item 2)								
Correct	45	57	50	52	60	58	59	69
Fallacy	29	27	35	35	32	34	31	30
Undefined	27	16	15	12	8	8	10	1
Expression (Item 2)								
Numerical	71	82	84	87	91	91	89	97
N	49	177	543	889	653	914	122	269

DISCUSSION

The results of this study indicated that incidence of conjunction errors was not associated with grade, nor with chance measurement developmental level. Sex differences favouring males were observed, but there were no improvements across cohorts; both of these findings are consistent with those of Watson and Moritz (1998). The failure to observe a decrease in the fallacy with grade, as found by Fischbein and Schnarch (1997), is most likely related to the simplified form of the question which reduced the fallacy level for all grades. It may also reflect the absence of specific reference to conjunction probabilities in the Australian mathematics curriculum (Australian Education Council, 1991) and hence lack of attention in the classroom.

Students from higher grades and of higher chance measurement developmental levels were more likely, however, to express probability for Item 1 in a numerical form, and consequently to exhibit higher rates of correct responses and fewer undefined responses. This may reflect the increasing emphasis on quantitative measurement of chance in the years of schooling. Not surprisingly, expressions used in responses to the probability item (Item 1) differed from those to the frequency item (Item 2). It is surprising, however, that incidence of the conjunction fallacy was not lower for the frequency form than for the probability form, as had been found by Tversky and Kahneman (1983). $Max(a,b)=c$ responses were more common than $min<c<max$ responses for Item 2, but not so for Item 1. Further investigation revealed that part (c) was evaluated as the arithmetic mean of parts (a) and (b) for 65 responses to Item 1 and 103 responses to Item 2, whereas part (c) was evaluated as the sum of parts (a) and (b) for 30 responses to Item 1 and 153 responses to Item 2. These findings indicate that conjunction errors result from a variety of forms of reasoning by different students, not simply from averaging the two component parts to determine the conjunction (cf Tversky and Kahneman).

Several issues arise for the classroom from the results of this study. The generally disappointing performance indicating a relatively constant incidence of the conjunction fallacy in social contexts would indicate a need to address the topic more directly in the mathematics curriculum. That this suggestion involves the acknowledgement of making subjective probability estimates must be faced. If students are going to be making personal life decisions in risk-taking situations where various conditions and their conjunctions are involved, they need experiences discussing more than marbles in urns.

REFERENCES

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

Davidson, D. (1995). The representativeness heuristic and the conjunction fallacy effect in children's decision making. *Merrill Palmer Quarterly, 41*(3), 328-346.

Fantino, E., Kulik, J., Stolarz-Fantino, S., & Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses. *Psychonomic Bulletin & Review, 4*(1), 96-101.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*, 123-129.

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

Fisk, J. E., & Pidgeon, N. (1997). The conjunction fallacy: The case for the existence of competing heuristic strategies. *British Journal Of Psychology, 88*, 1-27.

Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin, 10*, 243-252.

Moritz, J. B., Watson, J. M., & Collis, K. F. (1996). Odds: Chance measurement in three contexts. In P. C. Clarkson (Ed.), *Technology in mathematics education* (pp. 390-397). Melbourne: Mathematics Education Research Group of Australasia.

Moritz, J. B. (1998). Long odds: Longitudinal development of student understanding of odds. In C. Kanes, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times* (pp. 373-380). Gold Coast: Mathematics Education Research Group of Australasia.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90*, 293-315.

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. Volume 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.

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgement procedures, including "signed summation". *Organizational Behaviour and Human Decision Processes, 37*, 230-253.

Acknowledgement

This research was funded by the Australian Research Council, Grant Numbers A79231392 and A79800950.