ASSESSING STUDENTS' DIFFICULTIES WITH CONDITIONAL PROBABILITY AND BAYESIAN REASONING

Carmen Díaz

Inmaculada de la Fuente

ABSTRACT. In this paper we first describe the process of building a questionnaire directed to globally assess formal understanding of conditional probability and the psychological biases related to this concept. We then present results from applying the questionnaire to a sample of 414 students, after they had been taught the topic. Finally, we use Factor Analysis to show that formal knowledge of conditional probability in these students was unrelated to the different biases in conditional probability reasoning. These biases also appeared unrelated in our sample. We conclude with some recommendations about how to improve the teaching of conditional probability.

KEYWORDS. Conditional Probability Reasoning, Assessment.

INTRODUCTION

Conditional probability and Bayesian reasoning are important for undergraduate students, since they intervene in the understanding of classical and Bayesian inference, linear regression and correlation models, multivariate analysis and other statistical procedures, which are often used in empirical research. Conditional probability reasoning is also a critical part of statistical literacy in addition to being highly relevant in education, psychology, medicine and other professional fields. Reasoning based on conditional probability appears in these areas in evaluation, decision-making, diagnosis, and making inferences from samples to populations.

In spite of this relevance and the fact that a previous study of literature showed us that there is a large amount of research on this topic, we found no comprehensive questionnaires to globally assess students' understanding and misconceptions on these topics. Current school curriculum documents stress the need for assessment instruments to support learning and provide information that will allow reliable and valid inferences to be made about students' understanding, regardless of the context of the assessment task (Callingham, 2006). Following this need, the aim of this research was to build a comprehensive instrument that could be used

to assess the different biases and misunderstanding related to conditional probability. A second goal was to explore our preliminary hypothesis that formal knowledge of conditional probability is not related to the biases described in the literature.

In this paper we first summarize previous research on conditional probability and then describe the building of the CPR (Conditional Probability Reasoning) questionnaire whose preliminary evaluation was presented in Díaz and de la Fuente (2006). We then explore possible relationships between items that assess formal knowledge and those assessing the biases described in the literature.

PREVIOUS RESEARCH ON CONDITIONAL REASONING

Research on the understanding of conditional probability has been carried out with both secondary school and university students. Fischbein and Gazit (1984) conducted teaching experiments with children in grades 5-7 (10-12 year-olds) and compared two types of problems: In "without replacement" experiments (e.g., item 4 in Appendix) an element is selected from a set and not replaced and then a second selection is done. In "with replacement" experiments the element is put again in the initial set before selecting a second element. Fischbein and Gazit found that conditional probability problems were harder for these children in "without replacement" situations as compared to "with replacement" problems. Following that research Tarr and Jones (1997) identified four levels of thinking about conditional probability and statistical independence in middle school students (9-13 year-olds) (a more detailed description is given in Tarr & Lannin, 2005):

- Level 1 (subjective): Students ignore given numerical information in making predictions; they use subjective reasoning in assessing conditional probability and independence.

- Level 2 (transitional): Students demonstrate some recognition of whether consecutive events are related or not; however, their use of numbers to determine conditional probability is inappropriate.

- Level 3 (informal quantitative): Students' differentiation of "with and without replacement situations" is imprecise as is the quantification of the corresponding probabilities; they are also unable to produce the complete composition of the sample space in judging independence.

- Level 4 (numerical): Students state the necessary conditions for two events to be related, they assign the correct numerical probabilities and they distinguish between dependent and independent events in "with replacement" (e.g. item 15 in appendix) and "without replacement" (items 4, 9) situations".

Even when students progress towards the upper level in this classification (see also Tarr & Lannin, 2005), difficulties still remain at high school and university. This is shown in the various studies we summarize below, from which we have taken some of the items in our questionnaire. The full questionnaire is included in the Appendix.

Conditioning and Causation

It is well known that if an event *B* is the cause of another event *A*, then whenever *B* is present *A* is also present and therefore *P(A/B)=1*. On the contrary, *P(A/B)=1* does not imply that *B* is a cause for *A*, though the existence of such a conditional relationship indicates a possible causal relationship. From a psychological point of view, the person who assesses the conditional probability *P(A/B)* may perceive different type of relationships between *A* and *B* depending on the context (Tversky and Kahneman, 1982a). If *B* is perceived as a cause of *A*, *P(A/B)* is viewed as a causal relation, and if *A* is perceived as a possible cause of *B*, *P(A/B)* is viewed as a diagnostic relation. At other times people confuse the two probabilities *P(A/B)* and *P(B/A)*; this confusion was termed the fallacy of the transposed conditional (Falk, 1986). Item 7 in the CPR was included to assess these difficulties.

Causal Reasoning and the Fallacy of the Time Axis

Falk (1989) gave item 9 in Appendix to 88 university students and found that while students easily answered part 1, in part 2 they typically argued that the result of the second draw could not influence the first, and claimed that the probability in part B is 1/2. Falk suggested that these students confused conditional and causal reasoning and termed *fallacy of the time axis* their belief that an event could not condition another event that occurs before it.

This is a false reasoning, because even though there is no causal relation from the second event to the first one, the information in the problem that the second ball is red has reduced the sample space for the first drawing. Hence, *P (W1 is red/ W2 is red) =1/3*. Similar results were found by Gras and Totohasina (1995) who identified two different misconceptions about conditional probability in a survey of seventy-five 17 to 18 year-old secondary school students:

- The *chronological conception* where students interpret the conditional probability *P(A/B)* as a temporal relationship; that is, the conditioning event *B* should always precede event *A*.

- The *causal conception* where students interpret the conditional probability *P(A/B)* as an implicit causal relationship; that is, the conditioning event *B* is the cause and *A* is the consequence.

Synchronic and Diachronic Situations

Another issue involving time and conditional probability has been identified in the literature. In *diachronic* situations (e.g. items 4, 8, 9 and 10 in Appendix) the problem is formulated as a series of sequential experiments, which are carried out over time. *Synchronic* situations (e.g. items 1, 5, 14 and 17 in Appendix) are static and do not incorporate an underlying sequence of experiments. Formally the two situations are equivalent, however Sánchez and Hernández (2003) in their investigation with one hundred and ninety-six 17 to 18 year-old students found that students did not always perceive the situations as equivalent. These students add probabilities instead of using the product rule when computing a compound probability in a synchronic problem but use the correct rule in a diachronic situation.

Solving Bayes Problems

Considering reasoning needed to solve problems involving the Bayes' theorem, early research by Tversky and Kahneman (1982a) suggested that people do not employ this reasoning intuitively (see a summary in Koehler, 1996). Their research established the robustness and wide extent of the *base-rate fallacy* in students and professionals (Bar-Hillel, 1983). Mathematical analyses of tasks used in this research and of students' responses also reveal that their complexity is often greater than that reported in psychological research. For example, Totohasina (1992) analyzed the intuitive strategies of 67 pre-university students after being introduced to probability and before being taught Bayes' theorem. Only 25% gave correct responses. Totohasina suggested that part of the difficulty in solving Bayes' problems is due to the representation chosen by the student to solve the problems and that the use of a two way table is an obstacle to perceiving the sequential nature of some problems, and therefore can lead students to confuse conditional and joint probability. Another difficulty is the need to invert A and B in P(A|B) (condition and conditionant) in the problems, since students frequently confuse the role of these two events in a conditional probability.

Recent research suggests that Bayesian computations are simpler when information is given in natural frequencies, instead of using probabilities, percentages, or relative frequencies (Cosmides & Tooby, 1996; Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995). The suggested reason is that natural frequencies (absolute frequencies) correspond to the format of information humans have encountered throughout their evolutionary development. In particular, Bayes problems transform to simple probability problems if the data are given in a format of absolute frequencies. Sedlmeier (1999) analyzes and summarizes recent teaching experiments carried out by psychologists that follow this approach and involve the use of computers. The results of these experiments suggest that statistical training is effective if students are taught to translate statistical tasks to a suitable format, that includes tree diagrams and absolute frequencies. Martignon and Wassner (2002) organized a teaching experiment where students were taught to

solve Bayes problems with the help of tree diagrams and absolute frequencies. Participants in their study achieved about 80% of success in these problems after instruction.

Other Difficulties in Conditional Probability

Other difficulties include problems in defining the conditioning event (Bar-Hillel & Falk, 1982) and misunderstanding of independence (Sánchez, 1996; Truran & Truran, 1997). People also have problems with compound probabilities. Kahneman and Tversky (1982a) coined the term *conjunction fallacy* for people's unawareness that a compound probability cannot be higher than the probability of each single event.

METHOD

The building of our questionnaire was based on a rigorous methodological process, which included the following steps:

1. *Definition of the variable*: In educational measurement (e.g. Millman & Greene, 1989) a distinction is made between constructs (unobservable psychological traits, such as *understanding of conditional probability*) and the variables (e.g. score in a questionnaire) we use to make inferences about the construct. In order to achieve objectivity in defining our variable, we decompose the construct "understanding conditional probability" into content units. These content units were selected on the basis of a content analysis of 19 text books used in the teaching of statistics to undergraduates. To select the books, the list of references recommended in statistics courses was requested from 31 different universities in Spain. All the textbooks recommended by at least 4 different universities were analyzed, after discarding some books in which conditional probability was not included. The conditional probability content in these textbooks was analyzed and the definitions, properties, relationships with other concepts and procedures were classified into a reduced number of content units by means of a systematic and objective identification. The list of content units identified from this analysis is included in Table 1.

2. *Constructing an item bank*: The aforementioned analysis was complemented with our review of previous research on conditional probability reasoning. This review that also served to compile a sample of *n=49* different items used in this research, some of which had been used by different authors. The topics of the item pool covered the range of content units defined in step 1. These items were translated into Spanish and reworded to make their format homogeneous and improve their understanding.

3. *Selection of items*: The item difficulty (percentage of correct responses) and discrimination (correlation with test total score) were estimated from the answers by different samples of students (between 49 and 117 students answered each pilot item). Selection of items to build

the pilot questionnaire took into account these two parameters as well as results from expert judgment. Ten statistics education researchers from five different countries (Brazil, Colombia, Mexico, Spain and Venezuela) who had themselves carried out research related to conditional probability or independence were asked to collaborate. They were asked to rate (in a 5-point scale) the adequacy of the content units to understanding conditional probability as well as the suitability of each item to assess understanding for each specific content unit. The items in the pilot questionnaire were selected in such a way that a) the intended content of the questionnaire was covered (see Table 1); b) there was an agreement from the experts about the item adequacy; and c) item difficulty and discrimination were suitable.

Table 1. Primary content assessed by each item in the CPR questionnaire

Content units included in the CPR questionnaire	CPR item
1. Defining conditional probability; giving appropriate examples	11
2. Recognizing that a conditional probability involves a restriction in the sample space	12
3. Base rate fallacy	2
4. Distinguishing conditional, simple and joint probabilities	1
5. Distinguishing a conditional probability and its inverse (transposed conditional fallacy	7
6. Conjunction fallacy	6
7. Distinguishing independent and mutually exclusive events	3
8. Computing conditional probabilities in a single experiment	13
9. Solving conditional probability problems in a sampling "with replacement" setting	15
10. Solving conditional probability problems in a sampling "without replacement" setting	4, 9
11. Computing conditional probabilities from joint and compound probabilities	5
12. Solving conditional probability problems when the time axis is reverted	8, 9
13. Distinguishing conditional, causal and diagnosis situations	7
14. Solving conditional probability problems in a diachronic setting	4.8, 9,10
15. Solving conditional probability problems in a synchronic setting	1,5,14,17
16. Solving compound probability problems by applying the product rule to independent events	16
17. Solving compound probability problems by applying the product rule to dependent events	10, 17
18. Solving total probability problems	14
19. Solving Bayes problems	2, 18

4. *Formatting and revising the item*: We included two different formats: a) Multiple choice items with 3-4 possible responses were used to allow quick evaluation in the sample of some of the most pervasive biases described in the previous literature, e.g. item 2, taken from Tversky and Kahneman (1982a) evaluates the base-rate fallacy, item 3 taken from Sánchez (1996) assesses the confusion between independent and mutually exclusive events and item

6 taken from Tversky and Kahneman (1982b) assesses the conjunction fallacy; b) Open-ended items were also used to better understand students' strategies in problem solving (e.g. item 18) and their understanding of definitions and properties (e.g., items 11, 12).

5. *Pilot trial of the instrument*: The pilot study took place in the academic year 2003-2004 with a small sample of *n=57* psychology major students in order to make a preliminary estimation of the questionnaire reliability and validity. A second sample of *n=37* students majoring in mathematics was used to compare the performances in the two groups and to identify items with and without discriminative properties.

6. *Revising the pilot questionnaire*. After discarding those items with bad psychometric features, there was a second revision of the questionnaire. Thirteen expert methodology instructors were given three alternative wordings for each item and asked to select the best version, considering methodology standards, as well as give the reasons for their choice. For each item the version preferred by the majority was selected and additional suggestions by the methodology instructors were used to improve readability further.

The final questionnaire (see an English translation in Appendix) is composed of 18 items, with some sub-items which score independently, and some open-ended items. Table 1 presents the items' primary contents that cover the content in the books analyzed as well as main biases described in the literature.

Sample

Students from the Universities of Granada (4 different groups of students; *n=308* students total) and Murcia (two different group of students; *n= 106* students total) comprised the sample (*n=414*). The students were enrolled in an introductory statistics course in the first year of University studies (typically, 18-19 year-olds). They had studied conditional probability at secondary school level and were taught conditional probability and the Bayes theorem with the help of tree diagrams, two-way tables and meaningful examples, for about 2 weeks before they completed the questionnaire. The questionnaire was given to the students as an activity in the course of data analysis.

RESULTS AND DISCUSSION

Reliability and Validity

Once the data were collected, we analysed the response of each student in each item, taking into account the completeness of response in the open-ended items. Students were given one point per each correct response in items 2-8, and correct response in each part of items 1 and 9. In items 10-17, they were given 1 point if the response was basically correct with minor

mistakes (e.g. in carrying out an arithmetic operation; see Table 2). In item 18 students were given a score ranging between 0-4 according the correctness of response (see Table 4).

The empirical distribution of scoring ranged between 3 and 30 with an average value of 19.12, a little higher than half the maximum possible score (33 points) and a standard deviation of 5.91. A first approach to the reliability of the instrument was carried out by computing the Alpha coefficient that gave a moderate value (Alpha=0.79), which is reasonable, given that the questionnaire was designed to assess a wide range of knowledge (see Table 1), so that a particular student might understand some of the concepts but not others. A second estimation of reliability using test-retest in a sub-sample of 106 students, who were given the questionnaire a second time a month later, provided a reliability coefficient of 0.91. Mean scores and variance in the two occasions were almost identical, and thus, a learning effect in the second time was unlikely. The theoretical analysis of the questionnaire content as well as the results from experts' judgment served to justify content validity, by comparing the content evaluated by each item to the content units included in the definition of the variable (Table 1).

Responses to Items

In computing several probabilities from a two-way table (item 1) 90% of the students correctly computed the simple probability, 61%, the joint probability and 59% and 56%, respectively the two conditional probabilities. This confirms Falk´s (1989) opinion that verbal ambiguity in linguistic expression of conditional probability makes it difficult for the student to distinguish conditional and joint probabilities, even after instruction.

Table 2. Percentages of the different responses in multiple-choice items of the CPR (n=414)

Item	a	b	c	d	Blank
			Response		
Item 2	8	7	29	50($^+$)	5
Item 3	28	15	29($^+$)	20	8
Item 4	1	89($^+$)	10	(*)	0
Item 5	35($^+$)	31	34	(*)	0
Item 6	25($^+$)	9	62	(*)	4
Item 7	6	32($^+$)	59	(*)	9
Item 8	77	9	10 ($^+$)	2	2
Item 9(1)	6	17	69 ($^+$)	7	1
Item 9(2)	24($^+$)	25	9	36	6
Item 10	9	13	76($^+$)	(*)	2

(*) Does not apply ($^+$)Correct

Results in Table 2 suggest the existence of the following reasoning conflicts among the students in the sample:

1. *As regards independence*: We found confusion of independence with mutual exclusiveness in 28 % of the responses to distracter (a) in item 3, a bias also noticed by Sánchez (1996). The chronological conception of independence described by Gras and Totohasina (1995) was also shown in 29% of the responses to distracter (b) in item 3.

2. *Concerning conditional probability*: 31% of the students confused conditional with a joint probability (response (b) in item 5) or with a simple probability (34% responses c in item 5). The conjunction fallacy was observed in 71% of the responses to item 6 and the confusion of the transposed conditional in 59% of the responses in item 7. Difficulties in computing probabilities when the time axis is inverted are suggested by the responses to items 8 and 9(2), although the chronological conception of conditional probability described by Gras and Totohasina (1995) was not so clearly shown in these two items.

The base rate fallacy was not so pervasive as suggested in previous research (Bar-Hillel, 1983) as shown in the responses to distracters (a) and (b) in item 2: The majority of students gave the correct response (d) in this item, thus suggesting improvement of base rate with instruction. Item 4 (computing conditional probabilities in a "without-replacement" setting) was also very easy.

Considering responses in open-ended items, results in Table 3 suggest that students had difficulties in giving a sound definition and an example of conditional probability (item 11) but were conscious of the restriction of sample space (item 12). They had difficulties in solving a conditional probability problem in a single experiment (item 13) due to a lack of distinction of dependent and independent experiments in the context (synchronic situation; l), so that many of them did not appear to have completely reached Level 4 in the conditional probability reasoning scheme by Tarr and Jones (1997). Solving total probability (item 14) and solving conditional probability problems "with replacement" problems (item 15) was easier than computing compound probability in the case of independent (item 16) and dependent (item 17) events.

Table 3. Percentage of solutions in open-ended items (n=414)

	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17
Blank or totally wrong	29	15	47	18	21	30	24
Partly correct	30	21	18	21	9	18	16
Correct solution	41	64	35	61	70	52	60

In solving an open-ended Bayes problem (item 18, see Table 4), more than half the students were able to compute the total probability and a little less gave the complete solution; the majority was at least capable of correctly identifying the data and even identifying the

probability to be computed although 16% failed in developing the total probability formula. We remark that data were given in the percentage format, which is considered harder than absolute frequency formats in Gigerenzer (1994) and Gigerenzer and Hoffrage´s (1995) research.

Table 4. Completeness of solutions in solving a Bayes problem (Item 18)

	Percentage
Blank or totally wrong	16
Correct identification of data	15
Identifies the inverse conditional probability,	16
Correct computation of denominator (total probability)	7
Correct solution	46

We conclude that, in general, the instruction was successful as regards problem solving capabilities, whenever there was no psychological bias involved in the situation. However, some of the biases described in the literature seemed not to be overcome with instruction.

Structure of Responses

To explore our conjecture that biases of conditional probability reasoning are unrelated to mathematical performance in the tasks, we carried out a factor analysis of the set of responses to all the items using the SPSS software. As we have described, students were given a score in the range (0-1) in items 2-8 and in each part of items 1 and 9, a score (0-2) in items 10-17 and a score (0-4) in item 18, according to the correctness of response. Before carrying out the factor analysis we standardized all the variables, so that all of them had the same contribution to the analysis. The factor extraction method was principal components, which is the most conservative method, as it does not distort the data structure. In Table 5 we present the factor loadings (correlations) of items with the different factors after Varimax rotation (orthogonal rotation; maximizing variance of the original variable space).

We found seven factors with eigenvalues higher than 1 that explained the following percentages of the total variance: 21% (the first factor), 7 % (the second factor), and about 6% in each of the remaining factors. A total of 59% of the variance was explained by the set of factors, which suggests the specificity of each item, and multidimensional character of the construct, even when there is a common part shared by all of the items.

These percentages of variance also revealed the greater importance of the first factor, to which most of the open- ended problems contribute, in particular solving Bayes' problems had the higher contribution, followed by solving total probability and compound probability problems. All of these problems require a solving process with at least two stages, in the first of

which a conditional probability is computed, which is used in subsequent steps (e.g. product rule). We interpret this factor as *solving complex conditional probability problems ability*.

Table 5. Factor Loadings for Rotated Components in Exploratory Factor Analysis of Responses to Items (only values >.3 are shown)

Item	1	2	3	4	5	6	7
Item 18. Bayes	.77						
Item 14. Total probability	.76						
Item 17. Product rule, dependence, synchronic	.76						
Item 16. Product rule, independence	.67						
Item 15. Conditional probability, "with replacement"	.43		.42				
Item 12. Sample space	.40		.46				
Item 2. Base rates/ Bayes	.34						.48
Item 1b. Conditional probability, table	.32	.61					
Item 1c. Joint probability, table		.79					
Item 1a. Simple probability, table		.61					
Item 1d. Conditional probability. table		.77					
Item 13. Conditional probability, single experiment			.67				
Item 11. Definition			.59				
Item 4. Conditional probability, "without replacement", diachronic			.39		.44		
Item 5. Conditional probability. from joint and compound probability, synchronic					.66		
Item 6. Conjunction fallacy					.62		
Item 7. Transposed conditional /causal-diagnostic						-.65	
Item 9b. Time axis fallacy, diachronic				.71			
Item 8. Time axis fallacy, diachronic				.70			
Item 10. Product rule dependence, diachronic				.35			-.46
Item 9a. Conditional probability, "without replacement"						.66	
Item 3. Independence /mutually exclusiveness							.68

*Computing simple, joint and conditional probability from a two-way tabl*e (item 1) appeared as a separate factor, probably because the task format affected performance, a fact which has also been noticed by Ojeda (1996) and Gigerenzer (1994), among other researchers. A third factor showed the relationships between *definition, sample space and computation of*

conditional probabilities in "with and without replacement" situations. These relationships suggest that the third factor requires Level 4 reasoning in Tarr and Jones (1997) classification.

The remaining factors suggested that the different biases affecting conditional probability reasoning that are described in this paper, appeared unrelated to mathematical performance in problem solving understanding (Factor 1), computing conditional probability from a two-way table (Factor 2), and to Tarr and Jones's (1997) level 4 reasoning (Factor 3) (as the related items were not included in the three first factors). Each of the biases (transposed conditional, time axis fallacy, conjunction fallacy, independence/mutually exclusiveness/synchronic setting) also appeared unrelated to one another; in some cases some of them were opposed or related to some mathematical components of understanding conditional probability. For example, independence was linked to the base rate fallacy (where people have to judge whether the events are independent or not) and opposed to the idea of dependence.

IMPLICATIONS FOR TEACHING

In this research the students' performances in the formal components of the test was quite good. In particular, we observed a high percentage of correct or partly correct solutions to problems (including total probability and Bayes problems). However, some of the biases described in the literature were widespread in these students' thinking.

The complex relationship between probabilistic concepts and intuition was also shown in the results of Factor Analysis where items assessing the biases in conditional probability reasoning were unrelated to those assessing formal knowledge. This complex relationship was also shown in the historical development of the topic, as described in Batanero, Henry and Parzysz (2005). Even when independence and conditional probability was informally used from the very beginning of the study of chance games, only in the middle of the 18th century these two concepts were made explicit in the mathematical theory. Furthermore, the formal modern definition of independence was criticized by von Mises (1928/1952), because even when this definition expanded the concept, it is not intuitive at all. It is natural that these historical difficulties recur in the students' learning of probability.

Consequently, our research suggests the need not only for reinforcing the study of conditional probability in teaching data analysis at university level but also for a change of approach in this teaching. As suggested by Feller (1968, p. 114) "the notion of conditional probability is a basic tool of probability theory, and it is unfortunate that its great simplicity is somewhat obscured by a singularly clumsy terminology". Following Nisbett and Ross' recommendations (1980), students should be "given greater motivation to attend closely to the nature of the inferential tasks that they perform and the quality of their performance" (p. 280) and consequently "statistics should be taught in conjunction with material on intuitive strategies

and inferential errors" (p. 281) of the sort presented in their book and in this paper. In this sense we support Rossman and Short (1995), who suggest conditional probability can be taught in line with new statistics education ideas, in presenting a variety of applications to realistic problems, proposing interactive activities and using valuable representations (such as two-way tables and tree diagrams), as well as technology to facilitate learning.

REFERENCES

Batanero, C., Henry, M., & Parzysz (2005). The nature of chance and Probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning*, (pp. 20-42). New York: Springer.

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), *Decision making under uncertainty*. (pp 39 – 61). Amsterdam: North Holland.

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. Cognition 11, 109-22.

Callingham, R. (2006). Assessing statistical literacy: a question of interpretation? In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. On line: http://www. stat.auckland.ac.nz/~iase.

Cosmides, L., & J. Tooby. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition,* 58, 1-73.

Díaz, C., & de la Fuente, I. (2006). Assessing psychology students' difficulties with conditional probability and bayesian reasoning. In A. Rossman & B. Chance (Eds.), P*roceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. On line: http://www. stat.auckland.ac.nz/~iase.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 249-267). New York: Cambridge University Press.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores* (Initial conceptions on statistical association and its evolution after a teaching experiment based on the use of computers). Unpublished Ph.D. University of Granada, Spain.

Falk, R. (1986). Conditional probabilities: insights and difficulties. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Sta¬tistics*. (pp. 292–297). Victoria, Canada: International Statistical Institute.

Falk, R. (1989). Inference under uncertainty via conditional probabilities. In R. Morris (Ed.), S*tudies in mathematics education: Vol.7. The teaching of statistics* (pp. 175-184). Paris: UNESCO.

Feller, W. (1968). An introduction to probability theory and its applications (Vol. 1). New York: Wiley.

Fischbein, E., & Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions? *Educational Studies in Mathematics*, 15, 1-24.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). In G. Wright, & P. Ayton (Eds.), *Subjective probability* (pp. 129 – 161). Chichester: Wiley.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats (pp. 129-161). *Psychological Review*, 102, 684 – 704.

Gras, R., & Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité (Chronology and causality, conceptions causing epistemological obstacles to the notion of probability). *Recherches en Didactique des Mathématiques*, 15(1), 49 – 95.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavior and Brain Sciences*, 19, 1-54.

Martignon, L., & Wassner, C. (2002). Teaching decision making and statistical thinking with natural frequencies. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics. Cape Town, South Africa*: International Statistical Institute and International Association for Statistical Education. Online: http://www.stat.auckland.ac.nz/~iase.

Millman, J., & Greene, J. (1989). The specification and development of test of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (pp. 335 – 366). London: Macmillan.

Mises, R. von (1952). *Probability, statistics and truth* (J. Neyman, O. Scholl,, & E. Rabinovitch, (Trans.). London: William Hodge and company. (Original work published 1928).

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments*. Englewood Cliffs, NJ: Prentice Hall.

Ojeda, A. M. (1996). Contextos, representaciones y la idea de probabilidad condicional (Contexts, representations and conditional probability). In F. Hitt (Ed.), *Investigaciones en matemáticas educativas* (pp.291-310). México: Grupo Editorial Iberoamericano.

Pollatsek, A., Well, A. D., Konold, C., & Hardiman, P. (1987). *Understanding conditional probabilities. Organization, Behavior and Human Decision Processes*, 40, 255 – 269.

Rossman, A., & Short, T. (1995). Conditional probability and education reform: Are they compatible? *Journal of Statistics Education*, 3(2). On line: http://www.amstat.org/publications/jse.

Sánchez, E. (1996). Dificultades en la comprensión del concepto de eventos independientes (Difficulties in understanding independent events). In F. Hitt (Ed.), *Investigaciones en Matemática Educativa* (pp. 389–404). México: Grupo Editorial Iberoamericano.

Sánchez, E., & Hernández, R. (2003). Variables de tarea en problemas asociados a la regla del producto en probabilidad (Task variables in product rule problems in probability). In E. Filloy (Coord.), *Matemática educativa, aspectos de la investigación actual* (pp. 295 –313). México: Fondo de Cultura Económica.

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implications*. Mahwah, NJ: Erlbaum.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, 9, 39-59.

Tarr, J. E., & Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp.216-238). New York: Springer.

Totohasina, A. (1992). *Méthode implicative en analyse de données et application á l'analyse de concetpions d'étudiants sur la notion de probabilité conditionnelle* (Implicative data analysis and application to analyse students' conceptions on conditional probability). Unpublished Ph.D. University of Rennes I.

Truran, J. M., & Truran, K. M. (1997). Statistical independence: One concept or two? In B. Phillips (Ed.), *Papers from Statistical Education Presented at Topic Group 9, ICME 8*, (pp. 87–100). Hawthorn, Victoria, Australia: Swinburne University of Technology.

Tversky, A., & Kahneman, D. (1982 a). Judgements of and by representativeness. In D. Kahneman P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84-98). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982 b). Evidential impact of base rates. In D. Kahneman P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). New York: Cambridge University Press.

APPENDIX: ENGLISH TRANSLATION OF THE CPR (CONDITIONAL PROBABILITY REASONING) QUESTIONNAIRE

Note: The CPR questionnaire was developed and applied in Spanish and have been translated to English to be included in this paper. The authors are happy to give permission to other researchers to use either this version or the Spanish version or to send the Spanish version to those requesting it.

Part I: Reading a table

Read the questions carefully and then reply to each question. Include all the numbers and operations you used to get the response.

Item 1. (Estepa, 1994)

In a medical centre a group of people were interviewed with the following results:

	55 years-old or younger	Older than 55	Total
Previous heart stroke	29	75	104
No previous heart stroke	401	275	676
Total	430	350	780

Suppose we select at random a person from this group:
   a.   What is the probability that the person had a heart stroke?
   b.   What is the probability that the person had a heart stroke and, at the same time is older than 55?
   c.   When the person is older than 55, what is the probability of having had a heart stroke?
   d.   When the person had a heart stroke, what is the probability of being older than 55?

Part II. Multiple-choice items

The following items consist of multiple choice questions. Read the questions carefully and then choose only a response.

Item 2 (Tversky & Kahneman, 1982a)

A witness sees a crime involving a taxi in a city. The witness says that the taxi is blue. It is known from previous research that witnesses are correct 80% of the time when making such statements. The police also know that 15% of the taxis in the city are blue, the other 85% being green. What is the probability that a blue taxi was involved in the crime?
   a.   80/100
   b.   15 /100
   c.   (15/100) X (80/100)
   d. $\dfrac{15 \times 80}{85 \times 20 + 15 \times 80}$

Item 3. (Sánchez, 1996)

A standard deck of playing cards has 52 cards. There are four suits (clubs, diamonds, hearts, and spades), each of which has thirteen numbered cards (2,..., 9, 10, Jack, Queen, King, Ace). We pick a card up at random. Let A be the event "getting diamonds" and B the event "getting a Queen". Are events A and B independent?

    a.   They are not independent, since there is the Queen of diamonds.

    b.   Only when we first get a card to see if it is a diamond, return the card to the pack and then get a second card to see if it is a Queen.

    c.   They are independent, since P(Queen of diamonds)= P(Queen) x P(diamonds).

    d.   They are not independent, since P(Queen /diamonds)   P(Queen).

Item 4.

There are four lamps in a box, two of which are defective. We pick up two lamps at random from the box, one after another, without replacement. Given that the first lamp was defective:

    a.   The second lamp is more likely to be defective

    b.   The second lamp is most likely to be correct.

    c.   The probabilities for the second lamp being either correct or defective are the same.

Item 5. Eddy (1982)

10.3 % of women in a given city have a positive mammogram. The probability that a woman in this city has both positive mammogram and a breast cancer is 0.8%- A mammogram given to a woman taken at random in this population was positive. What is the probability that she actually has breast cancer?

    a. $\dfrac{0.8}{10.3} = 0.0776$, 7.76%

    b. $10.3 \times 0.8 = 8.24$, 8.24%

    c.   0.8 %

Item 6. (Tversky & Kahneman, 1982 b)

Suppose a tennis player reaches the Roland Garros final in 2005. He has to win 3 out of 5 sets to win the final. Which of the following events are more likely or are they equally likely?

    a.   The player will win the first set

    b.   The player will win the first set but lose the match

    c.   Both events a) and b) are equally likely

Item 7. (Pollatsek at al. 1987)

A cancer test was given to all the residents in a large city. A positive result was indicative of cancer and a negative result of no cancer. Which of the following results is more likely or are they equally likely?
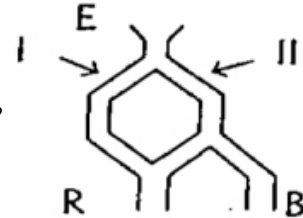
    a.   That a person had cancer if they got a positive result

    b.   Having a positive test if the person had cancer.

    c.   The two events are equally likely.

Item 8. Ojeda (1996)

We throw a ball in the entrance E of a machine (see the figure).

If the ball goes out through R, what is the probability of having passed by channel I?

    a.   1/2

    b.   1/3

    c.   2/3

    d.   Cannot be computed

Item 9. (Falk, 1986, 1989)

Two black marbles and two white marbles are put in an urn. We pick a white marble from the urn. Then, without putting the white marble in the urn again, we pick a second marble at random from the urn.

1. If the first marble is white, what is the probability that this second marble is white? P (W2/ W1)

    a.   1/ 2

    b.   1/ 6

    c.   1/ 3

    d.   1/ 4

2. If the second marble is white, what is the probability that the first marble is white? P (W1/ W2)

    a.   1/ 3

    b.   Cannot be computed

    c.   1/ 6;

    d.   1/ 2

Item 10.

An urn contains one blue marble and two red marbles. We pick up two marbles at random, one after the other without replacement. Which of the events below is more likely or are they equally likely?

    a.   Getting two red marbles.

    b.   The first marble is red and the second is blue

    c.   The two events a) and b) are equally likely.

Part II. Open-ended items

In the following items, please explain your reasoning in detail. Include all the numbers and operations you used.

Item 11. Explain in your own words what a simple and a conditional probability is and provide an example for each.

Item 12. Complete the sample space in the following random experiments:
   a) Observing gender (male/female) of the children in a three children family (e.g. MFM,...)
   b) Observing gender (male/female) of the children in a three children family when two or more children are male.

Item 13.

In throwing two dice the product of the two numbers was 12. What is the probability that none of the two numbers is a six (we differentiate the order of numbers in the two dice).

Item 14.

60% of the population in a city are men and 40% women. 50% of the men and 35% of the women smoke. If we pick a person from the city at random, what is the probability that the person is a smoker?

Item 15.

A person throws a die and writes down the result (odd or even). It is a fair die (that is all the numbers are equally likely). These are the results after 15 throws:

Odd, even, even, odd, odd, even, odd, odd, odd, odd, even, even, odd, odd, odd

The person throws once more. What is the probability to get an odd number this time?

Item 16.

A group of students in a school take a mathematics test and an English test. 80% of the students pass the mathematics test and 70% of the students pass the English test. Assuming the two subjects score independently, what is the probability that a student passes both tests (mathematics and English)?

Item 17.

According to a recent survey, 91% of the population in a city usually lie and 36% of those usually lie about important matters. If we pick a person at random from this city, what is the probability that the person usually lies about important matters?

Item 18. Totohasina (1982)

Two machines M1 and M2 produce balls. Machine M1 produces 40 % and M2 60% of balls. 5% of the balls produced by M1 and 1% of those produced by M2 are defective. We take a ball at random and it is defective. What is the probability that that ball was produced by machine M1?

Author   : **Carmen Díaz**

E-mail   : carmen.diaz@dpsi.uhu.es

Address   : Facultad de Psicología, Campus de «El Carmen»

                 Avda. de las Fuerzas Armadas, 21071 Huelvaa

Phone   : 34 959219354

Fax   : 34 959219357

Authors   : **Inmaculada de la Fuente**

E-mail   : edfuente@ugr.es

Address   : Metodología de las Ciencias del Comportamiento

                 Facultad de Psicología, Campus de Cartuja, 18071 Granada, Spain

Phone   : 34 958246270

Fax   : 34 958243746