UNDERSTANDING STATISTICAL MISCONCEPTIONS

Herman Callaert
Limburgs Universitair Centrum
Belgium

This paper reports on a preliminary study conducted for gaining better insight in the complexity of students' misconceptions of representativeness. Data from 156 students (112 high school graduates and 44 students with a university degree) are presented. The overall outcome indicates a lack of ability to refer problems about specific experiments to their correct context. Some results seem to contradict part of the representativeness heuristic described by Kahneman and Tversky (1972). They might also indicate that multiple-choice tests, even with two-part questions, are not able to fully capture the deep complexity of students' misunderstandings.

INTRODUCTION

During the past decades, research on statistical literacy and statistics education has established itself as an important and rapidly growing research field. Publications are numerous, and they cover a wide variety of topics, such as assessment in statistics education, the role of technology, the training of researchers, and so on (e.g. Batanero, 2001; Gal & Garfield, 1997; Garfield & Burrill, 1997; Ottaviani, 1996). A fundamental research area, covered extensively in the literature, is related to "statistical reasoning". Efforts for gaining a better insight in the process of statistical reasoning often include the study of "statistical misconceptions". Different categories of misconceptions have been identified, and within each category many examples and counterexamples have been explored and analyzed. The current paper, reporting on research in progress, addresses one specific example out of the vast area of possible misconceptions. Therefore, reference is made only to a seminal paper by Kahneman and Tversky (1972), and to a recent paper in the Journal of Statistics Education (Hirsch & O'Donnell, 2001). Of course, literature in this field is abundant.

ITEMS

For this report, questionnaires were compiled from the following set of items. For different groups of students, the questionnaires contained a specific subset of items, administered in a specific order in time. Some items appear quite similar, but the (apparently small) differences in formulation are crucial in this type of research, as will be explained further in this paper.

Item 1. A fair coin is tossed six times and the results are recorded in the order they appear. At each toss, the coin lands either H (=heads) or T (=tails).

The outcome H T H T T H
☐ is less likely than
☐ is as likely as
☐ is more likely than
the outcome H H H H H H

Item 2. A fair coin is tossed six times [and lands either H (=heads) or T (=tails)]. An outcome that contains three heads and three tails
☐ is less likely than
☐ is as likely as
☐ is more likely than
an outcome that has six heads.

Item 3. A fair coin is tossed six times. At each toss, the coin lands either H (=heads) or T (=tails).

An outcome like H T H T T H
☐ is less likely than
☐ is as likely as
☐ is more likely than
an outcome like H H H H H H

SAMPLE

For the current preliminary study, a sample of convenience has been used. A deeper analysis of the findings described in this report will be carried out, based on a precise sampling scheme and a revision of items in the questionnaires.

Different sets of questions were administered to students of different intellectual level. A first group (group A, *n*=14) consisted of students already holding a university degree (either bachelor or master) and who were just starting an MSc. program in applied statistics. All of those students have taken at least a couple of courses (at university level) in statistics during their previous study. At an even more sophisticated level, we have group B (n=30) of students entering a specialized MSc. program in statistics. These students all hold a master of science degree, mainly in mathematics or statistics. At a lower level, we have students who just finished high school and started university studies. Group C consists of students who took a lot of math/science courses in high school, also including topics in probability and statistics. Those students can be further subdivided into a very strong group of selected students (n=23) entering medical studies (group C1), and a group (n=35) starting in the biomedical field (group C2). A last group of students (group D, n=54) also just started university studies, but in the field of applied economics. It is known that several of those students do not take that many math/science courses in high school.

RESULTS

The current empirical research tries to gain insight into the heuristic called "representativeness" by Kahneman and Tversky (1972) (henceforth abbreviated K&T). K&T claim that, according to this heuristic, the subjective probability of an event, or a sample, is determined by the degree to which it: (i) is similar in essential characteristics to its parent population; and (ii) reflects the salient features of the process by which it is generated.

A standard question, studied over and over again in numerous research papers, deals with sequences of binary outcomes, such as boys and girls (K&T), heads and tails (Hirsch & O'Donnell, 2001), and so on. Classical mistakes made by students are related a.o. to the "ordering fallacy". K&T claim that this mistake occurs because students want the sample to be similar to the parent population. But what is the parent population of an ordered outcome of six tosses of a coin? Can this outcome be described as a simple random sample with replacement from a box containing two identical tickets, one with H written on it and one with T? If a student is reasoning along these lines, then, wrongly identifying HTHTTH as having higher probability than HHHHHH could be attributed to a misconception of representativeness. However, the student could be convinced that the parent population for samples yielding ordered n-tuples is itself an (ordered) n-dimensional random vector. For our example, this could be conceived as randomly drawing from a box containing 2^6 identical tickets. Why the HTHTTH ticket then should have a higher probability of being drawn than the HHHHHH ticket is not clear at all, but it certainly is not easily explained by a misconception of representativeness. It might be a misconception about the parent population.

In trying to discover students' reasoning and (possible) misconceptions of representativeness, we first addressed a group of well-trained university graduates (group A). They were given a questionnaire containing only item 1. One week later, they were asked to answer item 3. The results were quite amazing. As could be assumed, those students were aware of the "order fallacy" of item 1, having studied probability and statistics at university level. Indeed, the majority (11 out of the 14) correctly answered the question, in which is stated explicitly that order is important. Item 3 however is different in the sense that no mention is made of any ordering in the outcome. Indeed, it only asks for outcomes "like" some given result, giving the student plenty of opportunity for considering all "similar patterns". A large group of students however completely missed this point, and stuck to the answer they had given on item 1 the week before (see Table 1.)

Further investigation was undertaken with the students of group B. They were asked to answer two questionnaires, at the same occasion. The first questionnaire was the same as for group A (item 1). After having handed it in, they immediately received a second questionnaire, with item 2 and item 3 on the same page, and in that order. The intention was to make the expression “an outcome like” in item 3 very explicit, by first having them answer item 2, which is about unordered outcomes, formulated in a very precise way. Again, those students apparently knew about the order fallacy of item 1, and 24 out of 30 gave the correct answer to that item. But for item 2, where the probability of having three heads and three tails is twenty times higher than the probability of six heads, 18 out of 30 gave the erroneous answer that the outcomes are equally likely, and 2 students even marked the “less likely than” box (see Table 1). For item 3, they overwhelmingly marked “as likely as” (27 out of 30).

Table 1
Responses of Students with a University Degree

		Group A				Group B			
		responses to item 3				responses to item 2			
	responses	<	=	>	total	<	=	>	total
to item 1	<						1		1
	=		9	2	11	1	16	7	24
	>			3	3	1	1	3	5
	total		9	5	n=14	2	18	10	n=30

A similar study was carried out for students who just finished high school. Group C1 were students with a strong background in mathematics and sciences and who were admitted to medical studies after a special selection procedure. Their responses (see Table 2) show a pattern, which is remarkably close to (and is even better than) that of the students holding a university degree. Almost all of them correctly answered item 1, indicating that these students have a very good understanding of the difference between an ordered and an unordered sample. They certainly didn’t fall into the classical trap of referring to “essential characteristics of an underlying {H,T} population”. As for item 2, students in the group C2 perform a bit better on item 2 and a bit worse on item 1 than the university graduates of group B (Table 1 and Table 2). Finally, the test was administered to students with different high school background (group D) (see Table 3). They all just started university studies in the field of applied economics.

Table 2
Responses of High School Graduates with Strong Background in Math/Sciences

		Group C1				Group C2			
		responses to item 2				responses to item 2			
	responses	<	=	>	total	<	=	>	total
to item 1	<							1	1
	=		13	8	21		12	8	20
	>			2	2		6	8	14
	total		13	10	n=23		18	17	n=35

Table 3
Responses of High School Graduates with a Diverse Background in Math/Sciences

		Group D			
		responses to item 2			
	responses	<	=	>	total
to item 1	<		1		1
	=		24	5	29
	>	1	11	12	24
	total	1	36	17	n=54

DISCUSSION

As already mentioned, this work is a preliminary study only, and hence no formal statistical analysis is carried out at this point in time. Although some sample sizes are small, we nevertheless produce the following Table 4 with percentages, just for the sake of comparison.

Table 4
Percentage of Correct Responses

	group B	group C1	group C2	group D
item 1 correct	80 %	91 %	57 %	54 %
item 2 correct	30 %	43 %	49 %	31 %
both items correct	23 %	35 %	23 %	9 %

In his reflection on teaching probability and statistics, Shaughnessy (1992) emphasizes the need to (i) know more about how students think about probability, (ii) identify effective methods of instruction, and (iii) develop reliable methods of assessment that more accurately reflect students' conceptual understanding. That this is no easy task is reflected in the current study. Referring to the K&T paper, it is not clear what students consider to be the parent population of an event or a sample. Should effective methods of instruction then aim at "correct sampling schemes from the right underlying population" or at representativeness fallacies, or at the combination of both? Not one group of students succeeded in having half of them produce the right answer on item 2 (see Table 4). This looks like a contradiction to the representativeness misconception as described by K&T. Although a test with two-part multiple choice items as constructed by Hirsch and O'Donnell might give additional insight in students' misconceptions of representativeness, more complex misunderstandings may still be hiding underneath (see bottom line of Table 4).

It is well known that the difference between an ordered outcome (H,T,H,T,T,H) and an unordered outcome {H,T,H,T,T,H} is often overlooked by students. Indeed, the difference is technical, and in mathematics two different symbols (,) and { , } are used. Hence, a mathematically trained student might pay closer attention than his/her colleague in humanities when a word like "ordered" appears in a description of the outcome of an experiment. Mathematically trained students (like group A, B and C1) have encountered problems like item 1 before. In the current study they overwhelmingly gave a correct answer to item 1. This could be an indication that they do not have a misconception of representativeness. But what do they really understand and how broad is the context of their understanding (like: drawing with or without replacement, having ordered or unordered outcomes, conditional probability, independence, sample size, and so on)? Their answer on item 2 is very revealing and very disappointing too (see Table 4). Combining the correct results on both item 1 and item 2, the groups range from 35% down to 9%. In order to design instructional interventions created specifically to eliminate students' misconceptions, a deeper understanding of those misconceptions is certainly needed.

REFERENCES

Batanero, C. (Ed.). (2001). *Training researchers in the use of statistics.* Amsterdam: IASE and The International Statistical Institute.

Gal, I., & Garfield, J.B. (Eds.) (1997). *The assessment challenge in statistics education.* Amsterdam: IOS Press and The International Statistical Institute.

Garfield, J.B., & Burrill, G. (Eds.). (1997). *Research on the role of technology in teaching and learning statistics*. Amsterdam: The International Statistical Institute.

Hirsch, L.S., & O'Donnell, A.M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education* [Online], *9(2).*

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

Ottaviani, M.G. (Ed.) (1996). *Multimedia and new educational environments: Statisticians and experts exchange views.* Rome: Dipartimento di Statistica, Probabilita e Statistiche Applicate, Universita di Roma "La Sapienza".

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D Grouws (Ed.), *Handbook of research for mathematics* (pp. 115-147). New York: Macmillan.