PEOPLE'S INTUITIONS ABOUT RANDOMNESS AND PROBABILITY: AN EMPIRICAL STUDY

MARIE-PAULE LECOUTRE
ERIS, Université de Rouen
marie-paule.lecoutre@univ-rouen.fr

KATIA ROVIRA
Laboratoire Psy.Co, Université de Rouen
katia.rovira@univ-rouen.fr

BRUNO LECOUTRE
ERIS, C.N.R.S. et Université de Rouen
bruno.lecoutre@univ-rouen.fr

JACQUES POITEVINEAU
ERIS, Université de Paris 6 et Ministère de la Culture
poitevin@ccr.jussieu.fr

ABSTRACT

What people mean by randomness should be taken into account when teaching statistical inference. This experiment explored subjective beliefs about randomness and probability through two successive tasks. Subjects were asked to categorize 16 familiar items: 8 real items from everyday life experiences, and 8 stochastic items involving a repeatable process. Three groups of subjects differing according to their background knowledge of probability theory were compared. An important finding is that the arguments used to judge if an event is random and those to judge if it is not random appear to be of different natures. While the concept of probability has been introduced to formalize randomness, a majority of individuals appeared to consider probability as a primary concept.

Keywords: *Statistics education research; Probability; Randomness; Bayesian Inference*

1. INTRODUCTION

In recent years Bayesian statistical practice has considerably evolved. Nowadays, the frequentist approach is increasingly challenged among scientists by the Bayesian proponents (see e.g., D'Agostini, 2000; Lecoutre, Lecoutre & Poitevineau, 2001; Jaynes, 2003). In applied statistics, "objective Bayesian techniques" (Berger, 2004) are now a promising alternative to the traditional frequentist statistical inference procedures (significance tests and confidence intervals). Subjective Bayesian analysis also has a role to play in scientific investigations (see e.g., Kadane, 1996). Formal applications of Bayesian probabilities are also more and more common in everyday life situations. Let us mention for instance the Bayesian spam-filtering techniques and the "probability of

Statistics Education Research Journal, 5(1), 20-35, http://www.stat.auckland.ac.nz/serj

precipitation" given by Canada's weather service (explicitly defined by that organization as a *subjective* estimate of rain or snow).

This change seriously questions the common choice in mathematics education to emphasize the teaching of the "frequentist" conception of probability and statistics and to virtually ignore the alternative Bayesian conception. Teaching the Bayesian approach appears nowadays both desirable and feasible (Berry, 1997; Lecoutre, Lecoutre, Grouin, 2001; Albert, 2002; Lecoutre, 2006), as previous objections (e.g., Moore, 1997) are less and less defendable. This should invite us not to radicalize the opposition between the Bayesian and frequentist inferences but rather to consider their interplay (see Bayarri & Berger, 2004). However, this requires a change of emphasis in the role of probability and randomness.

A recent empirical study indicates that students in introductory statistics class are generally confused about the different notions of probability (Albert, 2003). Clearly, continuing to teach *only* the frequentist conception cannot reduce the confusion. This implies to students either that there is only one "correct" conception of probability or that the frequentist and Bayesian conceptions are competitive, which should not be the case (Vranas, 2001). Moreover, an exclusive focus on frequentist notions may conflict with the students' intuitions and representations about probability (see e.g., Hawkins & Kapadia, 1984). In any case, as emphasized by Konold (1991, p. 144), "the teacher cannot, by decree, enforce a normative view."

We assume that variants of the concept of randomness are at the heart of probabilistic and statistical reasoning. In frequentist inference, only the data are random. As in the prototypical problems used in the traditional teaching of probability (flipping a coin, drawing a chip from a jar…), the frequentist conception involves a sequence of repeated trials or an ensemble of "identically" prepared systems. There is always a well-defined reference set of cases. So this seems to make probability an "objective" property of the data (of the coin, the chip…) existing in nature independently of us. In this sense, the frequentist approach emphasizes an "observable randomness" that can be "produced" (*simulated*). Unfortunately, *empirical* frequencies are seldom available for the assignment of probabilities in real problems. As a result, assigning a frequentist probability to a single case event is not often easy, since it requires *imagining* a reference set of events or a series of repeated experiments. Considerable teaching difficulties with the frequentist inference come from the fact that data are considered as random *even after observation*.

In the Bayesian approach, the parameters are also considered as random, while data after observation are fixed quantities. We need to use another conception of probability. The Bayesian probability is the *degree of belief* (or *confidence*) in the occurrence of an event or a measure of the degree of plausibility of a statement. It can serve to describe "objective knowledge," in particular based on symmetry arguments or on frequency data. It can also be used to express a *personal* description of a state of knowledge, eventually incorporating *subjective* opinions (Savage, 1954; de Finetti, 1974), a notion that the frequentist conception rejects as being problematic. With the Bayesian approach it is not *conceptually* problematic to assign a probability to a single case event. Moreover, the Bayesian definition fits the meaning of the term probability in everyday language, and so the Bayesian probability theory appears to be much more closely related to how people intuitively reason in the presence of uncertainty.

The calculus of probability has been introduced to formalize randomness. In the XIXth century, in accord with Laplace's determinist conception of the world, randomness is the word given to the ignorance of a person in a determined universe (Laplace, 1951). Nature is knowable and yields to mathematical rules. Randomness is either euphemism for ignorance, or the expression of the limits of human perception and knowledge; it is *randomness when unknown*. An alternative conception of randomness, that has been expressed as a fundamental principle in quantum physics, is that physical reality is irreducibly random; it is *randomness per se*.

The frequentist view of probability is often perceived as related to determinism. So, in his preface to the re-edition of Laplace's book on probability, Thom (1986, p. 8) wrote "Laplace est ouvertement 'fréquentiste' (Laplace is openly 'frequentist'), *comme il se doit* de quelqu'un qui postule le déterminisme universel [as someone who postulates universal determinism *must be*)" [italics added]. However, things are not so simple and Laplace is also often associated with the development of Bayesian ideas. One must admit that the concept of randomness is ambiguous and complex (Kac, 1983) and gives rise to various interpretations.

Consequently, what people — students as well as instructors — mean by randomness should be taken into account when teaching these topics. Since the early 1950s, psychologists have carried out extensive research on people's ability to *produce* or *perceive* randomness. In the first case, subjects were required to simulate a series of outcomes of some typical random process such as tossing a coin (for a review, see especially Wagenaar, 1972). In the second case, subjects were asked to rate the degree of randomness of several sequences of stimuli. One of the main conclusions of all these studies is that humans are not good at either *producing* or *perceiving* randomness (Falk & Konold, 1997; Nickerson, 2002). However, all these studies strongly involved a frequentist conception of probability. This is also the case for the numerous studies using simulations of sampling distributions in order to improve students' statistical thinking processes.

So much research remains to be done to inform the teaching of Bayesian statistics. Konold et al. (1991) postulated that whether what students think is random, or not random, had a role in understanding probability distributions. In this perspective, they carried out an exploratory study on people's subjective criteria of randomness. Twenty psychology students and five mathematicians were asked to categorize familiar items as either "random" or "not random." The authors distinguished two types of items. "Stochastic" items either involve a repeatable process (e.g., rolling a die) or consist of outcomes produced *via* a mechanism associated with chance (e.g., drawing from a set of objects); by contrast "real" items consist of outcomes defined from everyday life experiences (e.g., the germination of a planted seed). The study found the following results: (1) A higher percentage of stochastic rather than real items was classified as random by both students and mathematicians; (2) The subjects' justifications showed a great diversity of conceptions; (3) Some mathematicians expressed their difficulty with having to dichotomize the items because they tended to view randomness as an entity that is present in degrees.

We assume that the spontaneous criteria for assessing randomness are linked to the theoretical definitions of probability. So, a stochastic item implicitly involves a unique (well-defined) reference set of cases, and consequently can be assigned to either a frequentist or Bayesian probability about which it can be expected that different individuals agree. On the contrary, a real item describes a single case event for which a well-defined reference set does not exist a priori. Consequently, an individual who doesn't accept Bayesian probabilities may consider that it is impossible to assign a probability to a real item, either because he/she has no reference set (a frequentist probability doesn't exist) or because he/she considers that any reference set should depend of his/her *personal* experience (an objective probability cannot be calculated). Consequently, greater variability can be expected for the real items. The psychological effect of considering a single case rather than a set of cases was termed the "power of the particular" by Kahneman (quoted in Griffin & Buelher, 1999). It seems to encompass at least two separate phenomena: (1) More empathy and other emotional reactions are aroused with the single case, because it is easier to imagine and identify with; (2) It also invites analysis by reasoning processes that are case-specific and deterministic, rather than statistical.

With the purpose of further investigations, we devised a two-phase experiment in which the second phase was similar to the Konold et al. (1991) procedure, hence a "constrained categorization": here, *randomness was an imposed criterion*. It was preceded by a "free categorization" task: here there were *no imposed criteria*. We assumed that this task permitted us to gather answers as spontaneous as possible which should partly reflect the subject's representations and beliefs about randomness.

We compared three groups varying in expertise in probability: lower secondary school pupils, psychology researchers, and mathematical researchers. Our main objective was to provide evidence of some internal coherence in probability judgments.

2. METHODS

2.1. SUBJECTS

Three groups of 20 subjects participated in the experiment.

(1) COL group: 20 pupils of the third class of a "collège" in Rouen (in France this corresponds to the last class of the lower secondary school) were chosen at random. They were of both sexes and aged 14-16 years old. They had not had a course in either statistics or in probability.

(2) PSY group: 20 psychology researchers from universities in Rouen and Paris, all with a PhD. They were recruited if they had some training in probability and applied statistics and had some practical experience processing experimental data.

(3) MAT group: 20 mathematics researchers from the university of Rouen, all with a PhD. They were recruited if they had training in probability theory and in mathematical statistics and had experience in teaching probability or statistics.

2.2. MATERIALS

The 16 items, reported in Table 1, were presented on individual cards. They are a priori categorized into four classes. Eight real items are events from everyday life experiences. In 4 items the subject is implied in the formulation by the use of the personal pronoun "you," while this is not the case in the other 4 items. Eight stochastic items either involve a repeatable process or consist of events produced via a mechanism which is associated with chance. Four items involve two equally likely, symmetric outcomes, while the 4 other items involve asymmetric outcomes.

2.3. PROCEDURE

The subjects carried out the task individually. They were told that they would be taking part in an experiment aimed at assessing their spontaneous judgments on various familiar situations. The item cards were randomly mixed and simultaneously visible. First, the subjects were asked to "put together the cards which go together," and thus to make piles. They were told that they could make as many piles as they wished and could take as much time as they wanted (*free categorization*). After they completed this task, they answered the following question: "Why did you put those cards together, and those ones, and so on?" Then the cards were mixed again and the subjects were asked to answer the following question: "For each card, do you think that there is randomness involved or not; explain why" (*constrained categorization*). The experiment lasted from 15 to 30 minutes.

Table 1. List of the 16 items

Real situations

With an implication of the subject in the formulation of the situation

A You meet a friend you have not seen for 10 years
B You win 10000F at the lottery
C You say the first thing that comes to your mind
D You will get the flu in the next month

Without any implication of the subject in the formulation of the situation

E A planted seed germinates or does not.
F The quotation of a stock at the Stock Exchange of Paris will go up more than 5% in the next three months
G It rained in Paris on March 15, 1936 or did not
H It will rain tomorrow in Paris

Stochastic situations

With symmetric outcomes

S An even number is obtained from a rolling of a die
T Heads is obtained from the toss of a fair coin
U Tails is obtained at the fifth flip of a fair coin that has landed with tails up on the previous four flips
V A white marble is drawn from a box that contains 10 black and 10 white marbles

With non-symmetric outcomes

W Two red chips are drawn from a box that contains 1 white chip and 2 red chips
X A pair of socks that match is obtained from a blind draw of two socks from a drawer in which there are two pairs of different socks
Y A lemon-flavoured sweet is drawn from a box that contains 20 orange-flavoured and 10 lemon-flavoured sweets
Z A white marble is drawn from a box that contains 10 black and 20 white marbles

3. RESULTS

3.1. FREE CATEGORIZATION

The categorizations were analyzed using the additive similarity trees (AST) model (Sattath & Tversky, 1977). This is a valuable alternative to multidimensional scaling (MDS), in which each object is represented by a point in a multidimensional (usually Euclidean) space. In AST the objects are represented by the external nodes of a tree. Roughly speaking, in the first step a topology is found such that a condition called the "four-point condition" (Buneman, 1974) is verified at best, in a certain sense. This condition, stronger than the triangle inequality, is characteristic of an additive tree. The distances *d* between any four points *x*, *y*, *z*, *t* of the tree satisfy the inequality $d(x,y)+d(z,t) \leq max[\,d(x,z)+d(y,t),\, d(x,t)+d(y,z)\,]$. Then arc-lengths are scaled so that the length of the path joining two nodes has a close fit to the similarity between the corresponding objects. The four-point condition is weaker than the ultrametric (or strong triangle) inequality that must be satisfied by ultrametric trees associated with hierarchical clustering. Consequently, additive trees are more likely to provide a faithful representation of proximity data than ultrametric trees. Pruzansky, Tversky, and Carroll (1982) reanalyzed proximity data sets from various published studies and concluded that MDS was more appropriate when the hypothesized structure of the objects was perceptual and that AST was more appropriate when it was conceptual.

From the categorizations made by the subjects, an overall distance matrix was obtained; the distance between two items was the percentage of subjects who classified these two items into separate categories (thus the possible maximal distance between two items was 100). This matrix was input to the computer program ADDTREE (in the version by Barthélemy & Guénoche, 1991) to produce the additive similarity trees for the items.

The theoretical additive similarity tree associated with the a priori classification of the items is shown in Figure 1. The observed trees within each of the three groups of subjects are shown in Figure 2.

Figure 1. Theoretical additive similarity tree associated with the a priori *classification of the items into four classes.*

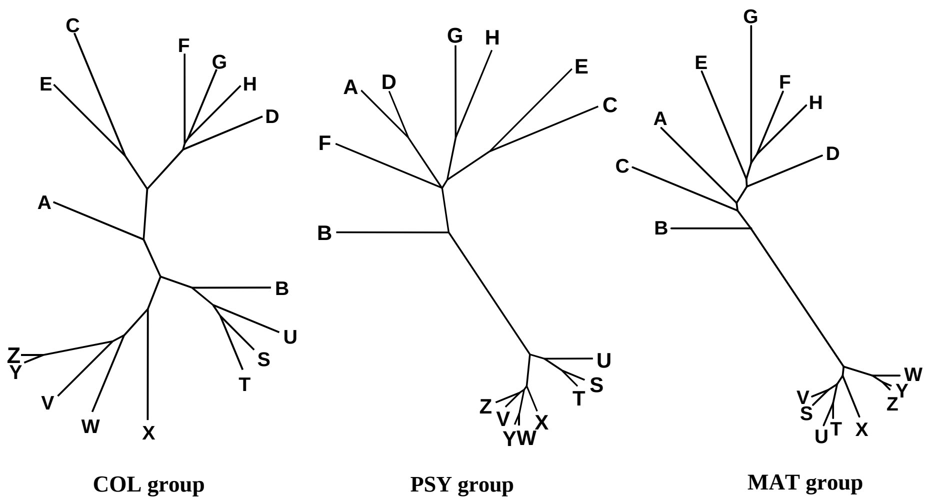


Figure 2. Free categorization: observed additive similarity trees within each of the three groups of subjects.

The tree-structures are quite similar for the three groups. The 16 items are first partitioned into two major clusters: real items versus stochastic items. The results are especially striking for the MAT group since all the real items (ABCEDFG) on the one hand and all the stochastic items (STUVWXYZ) on the other hand are regrouped at two extremities of the tree. Furthermore in the MAT group, each of these two clusters is further partitioned into two finer clusters. The eight real items are divided according to the degree of implication of the subject: EFGH (without any implication) are separated from ABC (with implication), D (flu) being more distant. The eight stochastic items are divided according to the nature of their outcomes: STUV (with symmetric outcomes) are separated from WXYZ (with non-symmetric outcomes). For the two other groups of subjects, although the same first partition into two major clusters can be observed, the partition into finer clusters is less apparent, and the trees are more widespread, especially for the COL group. It can be noted that item B (lottery) is closer to the stochastic items than to the real items for the PSY group, and is clearly associated with the stochastic items for the COL group.

The *d* distance of Robinson and Foulds (1981) between the trees is reported in Table 2. This distance is purely topologic, that is to say that it only takes into account the structure of the trees, while ignoring the length of the paths. It is equal to the minimum number of elementary operations (fusion or division of nodes) necessary to transform one tree into another one, and for *k* items lies between 0 and 2*k*-6, thus here $0 \leq d \leq 26$.

Table 2. The d distance of Robinson and Foulds ($0 \leq d \leq 26$)

	MAT group	PSY group	COL group
PSY group	22		
COL group	20	12	
Theoretical tree	12	16	18

It is for the MAT group that the distance between the theoretical tree and the observed tree is the smallest. So the free categorizations of the MAT subjects are those which fit the best with the a priori theoretical classification. The trees of the two other groups are nearly equidistant. Furthermore, it can be noted that the maximal distances (nearly equal) are observed between the tree of the MAT group and the trees of the other two groups.

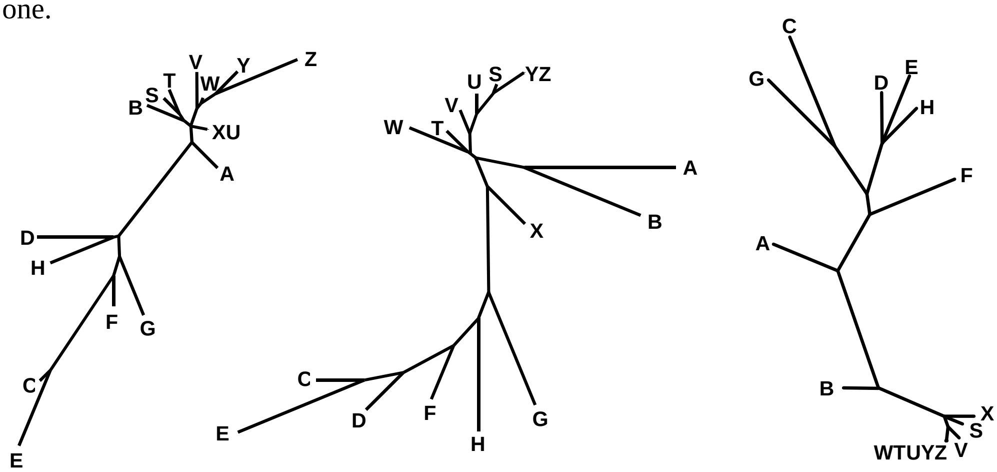
The justifications given by the subjects support further comments about the categorizations. There is large inter-individual variability, since there are almost as many different sets of categorizations and justifications as there are subjects. Nevertheless, a striking finding is that most subjects have explicitly used the notion of randomness in the free categorization task, although this term was never mentioned in the instructions. The main criterion used in all three groups involves the opposition between the items linked to a probability ("computable events") and those linked to everyday life experiences for which it is difficult if not impossible to calculate a probability. Other criteria are specific to each group. In the MAT group, we frequently observed an opposition between the events which are typical examples of standard models of randomness ("typical mathematical problems for our students") and the events in which there is randomness "when unknown" ("no available standard model"). Furthermore, within this group some categorizations are based either on the type of probability involved – conditional or elementary – or on the probability value (e.g., <1/2, 1/2, 1/3,…). These criteria can be viewed as variants of the aforementioned main criterion. In the two other groups there is greater inter-individual variability. In the PSY group, some subjects differentiated the events that are linked to nature or the environment (yielding to some meteorological or biological rules) from the "purely random" events. In the COL group, some categorizations are specifically based either on the opposition between *lucky* ("that's chance”) and *unlucky* (“that’s fatality”) events, or to the degree of implication of the subject.

3.2. CONSTRAINT CATEGORIZATION

A mathematician systematically answered “I don’t know” to all the items and was eliminated from the study. Some subjects, rather than expressing a dichotomous attitude, rated graduated judgments such as “randomness is involved, but only a little.” Consequently, the answers were a posteriori classified into three main categories: R (Random), L (a Little bit random) and N (Not random).

Trees The three additive trees for the items are reported in Figure 3. The tree-structures are quite similar for the three groups. On the whole the 16 items are partitioned into two major clusters: real items versus stochastic items. Nevertheless, there are two main exceptions concerning items A (friend) and B (lottery) which are separated from the other real items and are closer to the stochastic items.

A comparison of the trees obtained in the two phases, shows that making the notion of randomness explicit has three main effects: (1) For the stochastic items, the distinction between symmetric and non-symmetric outcomes is less apparent in the constrained categorization. So these two categories of items are perceived as quite similar when randomness is an explicit criterion of classification, while they can be perceived as different in the free categorization task; (2) In the constrained categorization the real items are much more dispersed, revealing more divergent conceptions when randomness is an explicit criterion; (3) The real items A (friend) and B (lottery) are relatively isolated and are much closer to the stochastic items in the constrained categorization than in the free one.



COL group **PSY group** **MAT group**

Figure 3. Constrained categorization: observed additive similarity trees within each of the three groups of subjects.

The justifications given by the subjects support further comments. For the stochastic items, the explicit reference to a random model in the constrained categorization is so salient that the distinction between symmetric and non-symmetric outcomes vanishes. For the two real items A and B in the free categorization, the presence/absence of implication is the salient property for most subjects in the three groups, which led them to classify these two items with other real items. In the constrained categorization, it is the extreme unusualness of these two events that became the salient property for most subjects. Consequently these two items are separated from the other real items and are closer to the stochastic items which are viewed in majority as events in which randomness intervenes.

Inferential analyses The percentages of answers within the three groups are reported in Table 3. For the real items, there is no consensus except for a certain extent A (friend) and B (lottery), which are mainly categorized as random, especially by the COL subjects. For the other real items, there is no large systematic majority. However, they are mainly categorized as not random by PSY subjects (from 55% to 75%). Items C (first thing) and E (seed) are the most frequently categorized as not random (by respectively 69% and 67% of the subjects).

For each stochastic item, there is a large majority of COL and MAT subjects (from 85% to 100%) who categorize it either as random or in a few cases as a little bit random. Nevertheless, not one item is unanimously categorized as random by all these subjects. By contrast, each of these items is categorized as not random by a non negligible and approximately constant proportion of PSY subjects (from 30% to 40%).

Table 3. Proportions of answers for each of the 16 items within the three groups of subjects: COL (n=20)*, PSY* (n=20) *and MAT* (n=19)

		Real items								Stochastic items							
		With implication				Without implication				Symmetric outcomes				Asymmetric outcomes			
		A	B	C	D	E	F	G	H	S	T	U	V	W	X	Y	Z
R	COL	.90	.90	.25	.50	.05	.50	.45	.50	.90	.90	1	.85	.95	1	.85	.70
	PSY	.60	.65	.20	.25	.15	.25	.25	.25	.50	.50	.45	.45	.60	.45	.45	.45
	MAT	.53	.79	.16	.37	.26	.32	.16	.32	.89	.95	.95	89	.95	.89	.95	.95
	mean	.68	.78	.20	.37	.17	.36	.29	.36	.76	.78	.80	.73	.83	.78	.75	.70
L	COL	.05			.15			.05	.15							.10	.20
	PSY	.10	.20	.05	.15	.10	.10	.15	.20	.20	.10	.20	.15	.10	.15	.25	.25
	MAT	.26	.05	.16	.37	.42	.32	.21	.42				.05				
	mean	.14	.08	.12	.22	.17	.14	.14	.26	.07	.03	.07	.07	.03	.05	.12	.15
N	COL	.05	.10	.75	.35	.95	.50	.50	.35	.10	.10		.15	.05		.05	.10
	PSY	.30	.15	.75	.60	.75	.65	.60	.55	.30	.40	.35	.40	.30	.40	.30	.30
	MAT	.21	.16	.68	.26	.32	.37	.63	.26	.11	.05	.05	.05	.05	.11	.05	.05
	mean	.19	.14	.69	.40	.67	.51	.58	.39	.17	.18	.13	.20	.13	.17	.13	.15

R: Random; L: a Little bit random; N: Not random

In all the further analyses, the two categories "Random" and "a Little bit random" were grouped together and coded R ("Random"). A contrast analysis was performed in a three-groups ANOVA design with a four-level repeated factor, "Items," corresponding to the four classes of items. The dependent variable was the proportions of items categorized as not random. Interval estimates are reported for each difference between proportions of interest. We used standard non-informative Bayesian procedures (see Lecoutre & Derzko, 2001) to assess either the largeness or the smallness of the population difference δ. Results are summarized in Table 4.

Table 4. Contrast analysis for the constrained categorization. The dependent variable is the proportions of items categorized as not random (N). *D is the observed difference and δ is the population difference.*

Contrast	*D*	*t* [df]	*p*	Bayesian conclusion
real versus stochastic	+.29	+7.06 [56]	<.0001	Pr(δ>+.24)=.90
real versus stochastic/COL	+.38	+6.38 [19]	<.0001	Pr(δ>+.30)=.90
real versus stochastic/PSY	+.20	+2.96 [19]	.0080	Pr(δ>+.11)=.90
real versus stochastic/MAT	+.30	+3.44 [18]	.0028	Pr(δ>+.18)=.90
implication versus no implication/real items	+.17	+4.78 [56]	<.0001	Pr(δ>+.13)=.90
symmetric versus asymmetric/stochastic items	+.03	+1.07 [56]	.2892	Pr(.01<δ<+.06)=.90
PSY *vs* COL+MAT	+.21	+2.96 [56]	.0044	Pr(δ>+.12)=.90

The observed difference between the unweighted average proportions of not random answers for the real (.45) and the stochastic items (.16) is *D*=+.29, significantly different from 0. The standard Bayesian analysis shows that there is a 90% probability that δ is larger than +.24; a notably large difference can be assessed. The same conclusions are found within each group of subjects (see Table 3).

Within the real items, the observed difference between the items without implication (.54) and those with implication (.36) is +.17 and a notably large difference can be assessed (Pr(δ>+.13)=.90). Within the stochastic items, the observed difference between the items with symmetric outcomes (.17) and those with asymmetric outcomes (.15) is +.03 (non significant) and a relatively small difference can be assessed (Pr(-.01<δ<+.06)=.90).

The observed difference between the average proportions of not random answers for the PSY group (.44) and the two other groups (.26 for the COL group and .21 for the MAT group) is +0.21. A notable difference can be assessed (Pr(δ>+.12)=.90) This difference is mainly attributable to the stochastic items that are more often categorized as not random within the PSY group than within the two other groups.

Individual patterns The individual patterns of the 16 answers given by each subject were further analyzed. Each pattern is a string of 16 Ns or Rs, ranked from items A to P. Taking into account the justifications given by the subjects, some general conceptions of randomness can be identified. Each identified conception defines a theoretical pattern. An example of a general conception is that randomness is involved whenever it is possible to calculate a probability; consequently, randomness is involved for stochastic items and is not involved for real items, hence the theoretical pattern NNNNNNNNRRRRRRRR.

We considered that an observed pattern was compatible with a theoretical pattern if at least 14 out of the 16 answers were the theoretical answers. This analysis shows that three general conceptions allow us to account for 75% (44/59) of the observed patterns.

(1) The majority conception is that randomness is involved whenever probabilistic reasoning is involved. Thus randomness is involved for *all items*. Thirty nine percent of the observed patterns (23/59) are compatible with a string of 16 Rs. This conception is more frequent in the MAT group (53%) than in the COL (35%) and PSY (30%) groups. Note that some subjects, especially in the MAT group, explicitly refer to two kinds of randomness: a “mathematical” randomness and a randomness “when unknown.” Mathematical randomness would be linked to the events for which it is possible to compute a probability (typically the stochastic items); randomness when unknown would

be linked to all the events for which there is no possibility to *easily* compute a probability (typically the real items). For these subjects randomness is involved in *all* items, but stochastic and real items differ as to the *nature* of randomness.

(2) Another frequently encountered conception is that randomness is involved for the stochastic items – because it is possible to compute probabilities – and is not involved for the real items (except A, friend, and B, lottery, for which it would also be involved) because determinism plays a great part and causal factors can be identified. Thirty one percent (18/59) of the observed patterns are compatible with the corresponding theoretical pattern (RRNNNNNNNRRRRRRR) which is an approximate dichotomy between real and stochastic items. This is the majority conception in the COL group (40% compared to 26% in each of the two other groups).

(3) A conception encountered only in the PSY group is “randomness is never involved.” Five percent (3/59) of the observed patterns are compatible with a string of 16 Ns. One of these three psychologists expressed a strong conviction that the world is entirely deterministic. The two other subjects stated that randomness is not involved whenever it is possible to compute a probability, or in their words “quantify.”

The 25% remaining patterns involve partial conceptions which correspond to some specific views of randomness and apply only to some items. We will mention three of these conceptions. (1) A phenomenon is random only when all the outcomes have the same probability (cf. the “equally-likely” justification in Konold et al., 1991). Consequently randomness would be involved in at least the four stochastic items with symmetric outcomes; a typical justification is “it’s pure random because we have 50/50 chances.” This can be compared with the “equiprobability-bias” (Lecoutre, 1992) according to which random events are thought to be equiprobable “by nature” or with the “uniformity belief” (Falk, 1992) according to which people have a strong intuitive tendency to assume equal probabilities for the various available options. (2) A phenomenon is random when there is no prior knowledge about the outcome, and thus no possibility to predict nor to control the result (cf. the “causality” and “uncertainty” justifications in Konold et al., 1991). For instance E (seed germination) is “not random because one can control the soil, the wetness…”, H (rain) is “not random, because there is a way of predicting the weather.” By contrast, items involving a die or a coin are “random because one can’t control or predict anything.” These justifications can be compared with the theory of Piaget and Inhelder (1951) according to which the emergence of the idea of chance is attributed to children’s realization of the impossibility of predicting oncoming events or of offering causal explanations. (3) Some justifications reflect a conception which connects the degree of intervention of randomness to the value of the probability. Randomness is said to be involved more as the probability decreases. For instance, W (chips) is “almost not random because the probability is relatively high;” by contrast B (lottery) “is really random, because the probability is very weak.”

4. CONCLUSION

Our study has confirmed that individuals hold a wide range of meanings for the concept of randomness, since in the two categorization tasks taken altogether there were as many different classifications as there were subjects, however simple and familiar the 16 items may be. These findings are in accordance with the results of many studies which have been taken as evidence that the concept of randomness leads to a lot of different interpretations, even by many who use it extensively in their work (Nickerson, 2002). Nevertheless it was possible to distinguish some general conceptions of randomness, and so to provide evidence of some internal coherence in probability judgments. The 16 items were partitioned into two main classes opposing the real and the stochastic items which were perceived as different. A large majority of individuals were in agreement for the stochastic items and categorized them as random because it is “easily” possible to compute a probability. In contrast individuals were divided for the real items; they categorized them either as random or not random with no large majority. Two main conceptions have been observed for the real items: Either randomness is involved because a probabilistic reasoning is involved, or randomness is not involved because determinism plays a great part or because causal factors can be identified. These findings are compatible with the “power of the particular” according to which the single cases “seem to invite analysis by reasoning processes that are case-specific and deterministic, rather than statistical.”

Another important finding was the little effect of background knowledge of probability theory on one’s views of randomness. In particular, the dichotomy of stochastic versus real items was observed within each of the three groups, including lower secondary school pupils without any background knowledge of probability theory. This is compatible with Konold’s conclusion according to which students have strong intuitions about probability and randomness prior to instruction (Konold, 1995).

However the PSY and MAT groups exhibited some distinctive features. Within the PSY group, each stochastic item was categorized as not random by about a third of the subjects, while within the two other groups all stochastic items were categorized as random by almost all the subjects. For some psychologists an item is not random whenever it is possible to compute a probability. We assume that this marginal conception could be linked to their statistical practice. Indeed psychologists routinely use null hypothesis significance tests and a common presentation of this procedure is that rejecting the null hypothesis implies rejecting randomness and consequently could justify deterministic conclusions about the data. So, Tryon (2001) wrote “rejection of the null hypothesis implies that the results are not due to chance and that therefore they must be both systematic and reproductible.” Furthermore psychologists categorized real items as not random more often than the other subjects.

A characteristic of the MAT group is that some subjects explicitly referred to two types of randomness: a “mathematical” randomness when it is easy to compute an objective probability (typically the stochastic items), and a randomness “when unknown” when it is not easy to compute a probability due to a lack of available standard probabilistic model (typically the real items).

Finally, it must be emphasized that the arguments used to judge if an event is random, and those to judge if it is not random, were found to be of different natures. In general subjects considered randomness to be involved in situations when probability was also involved, and considered randomness not to be involved when causal factors could be identified. To assess randomness, a large majority of subjects argued that “it is random because it is possible to compute a probability.” All these subjects applied this probability based argument to the stochastic items, and approximately half of them to the real items that are consequently judged as random. It is interesting to note that, while the concept of probability has been introduced to formalize randomness (“randomness implies probability”), a majority of individuals appear to consider probability as a primary concept (“probability implies randomness”). By contrast only a minority of subjects referred to more direct arguments such as “it is random because one can’t control or predict anything” (“no causality implies randomness”). To assess non randomness, the main argument is that “it is not random because there determinism plays a great part or because causal factors can be identified” (“causality implies non randomness”). Only a weak minority (two psychologists) used a probability based argument and surprisingly argued that "it is not random because it is possible to compute a probability" ("probability implies non randomness").

It must be acknowledged with Shaughnessy (1992, p. 468) that "the model of probability that we employ in a particular situation should be determined by the task we are asking our students to investigate, and by the types of problems we wish to solve." Conversely, the types of problems should be enlarged to go beyond "stylized situations" in which probability assignments are essentially based on considerations of symmetry. They should, in particular, include situations involving probability judgments and predictions about single case events. Of course, this is not an effortless avenue. Clearly, we must not restrict our attention to the conventional frequentist view of probability. The Bayesian conception allows the students to assign a probability to a wider range of situations (Steinberg & Von Harten, 1982). Moreover, the Bayesian definition can be applied to real life uncertainty situations in which it is often not possible to "easily" compute a probability.

Concentrating more specifically on the teaching of the Bayesian statistical inference approach, instructors face the difficulty of explaining to students that the parameters, as well as the statistics (*before observations*), are considered as random. According to our results, this difficulty should be all the more serious in that it is not easy to assign a probability. On the one hand, the sampling probability distributions of statistics clearly refer to stochastic situations. At least, in most familiar situations, sampling probabilities are relatively easy to compute, and the level of mathematical justifications can be adapted to the students. On the other hand, Bayesian probabilities about parameters refer to single case events. The elicitation of the prior probabilities is precisely one of the most often denounced difficulties with the Bayesian approach. A possible approach, advocated by Berry (1997), is to place emphasis on the fact that prior and posterior Bayesian distributions are subjective and to force students to assess their prior probabilities. However , this task is not easy for the students and Berry recognized that "they don't like it." An alternative strategy, based on our teaching experience with Bayesian methods (Lecoutre, Lecoutre & Grouin, 2001; Lecoutre, 2006), is to avoid – at least *in a first stage* – the issue of assessing a subjective prior distribution and to focus the teaching on "objective Bayesian analysis" (Berger, 2004), based on *noninformative* priors. Such priors fit well with the conception of *randomness when unknown*. As for the sampling probabilities, the resulting posterior probabilities are relatively easy to compute and the level of mathematical justifications can be adapted to the students. Once students have become familiarized with their use and interpretation, the introduction of "informative" prior distributions *at a later stage* is generally well-accepted.

A considerable difficulty in the teaching of the frequentist approach is that data continue to be treated as random *even after observation*. This seems so *strange* to students that the frequentist interpretation of confidence intervals hardly makes sense for them. However, according to our results, it is not so paradoxical that most statistical users erroneously interpret the frequentist confidence level as the (Bayesian) probability of the single event "the parameter lies between two fixed limits." Indeed, since a probability is available for this event, these users have no doubt that it is a random event. Furthermore, all attempts to teach the orthodox frequentist interpretation seems to be "a losing battle" (Freeman, 1993). Our suggestion is to replace, as much as possible, probabilistic formulations about sampling distributions with formulations in terms of "proportions of samples." Thus, the probabilistic formulations are mainly reserved for the Bayesian approach, minimizing a possible source of confusion. In conclusion, it remains a challenge for statistics educators to reduce students' confusions about the different notions of probability. In this perspective, it is important that they become familiar with the variety of meanings and beliefs about randomness. In particular, knowing what students think is, or is not, “random,” in relation to probability based arguments, should facilitate the communication between students and teachers regarding probability and statistical inference. Our finding that background knowledge of probability theory has little effect on one’s view of randomness implies that a mutual understanding is possible.

ACKNOWLEDGMENTS

Our special thanks go to Charlotte Détrie for improving our English. The remaining mistakes are ours.

REFERENCES

Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. In B. Phillips (Ed.), *Sixth International Conference on Teaching Statistics Proceedings.* Cape Town, South Africa [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute [Online: www.stat.auckland.ac.nz/~iase/publications/1/3f1_albe.pdf]

Albert, J. (2003). College students’ conceptions of probability. *The American Statistician 57*, 37-45.

D’Agostini, G. (2000). Role and meaning of subjective probability: some comments on common misconceptions. *AIP Conference Proceedings*, *568*. Melville.

Barthélemy, J.-P., & Guénoche, A. (1991). *Trees and proximity representations*. New York: Wiley.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, *19*, 58-80.

Berger, J. (2004). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*, 1-17.

Berry, D. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*, 241-246.

Buneman, P. (1974). A note on the metric properties of trees. *Journal of Combinatorial Theory (B)*, *17*, 48-50.

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition, 43*, 197-223.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*, 301-318.

de Finetti, B. (1974). *Theory of probability* (Vol.1). New York: Wiley & Sons.

Freeman, P. R. (1993). The role of *p*-values in analysing trial results. *Statistics in Medicine*, *12*, 1443-1452.

Griffin, D., & Buelher, R. (1999). Frequency, probability, and prediction: Easy solutions to cognitive illusions? *Cognitive Psychology*, *38*, 48-78.

Hawkins, A. S., & Kapadia, R. (1984). Children’s conceptions of probability – A psychological and pedagogical review. *Educational Studies in Mathematics, 15,* 349-377.

Jaynes, E. T. (2003). *Probability theory: The logic of science* (Edited by G.L. Bretthorst). Cambridge: Cambridge University Press.

Kac, M. (1983). Marginalia: What is random? *American Scientist*, *71*, 405-406.

Kadane, J. B. (1996). *Bayesian methods and ethics in a clinical trial design*. New York: John Wiley & Sons.

Konold, C. (1991), Understanding students' beliefs about probability. In E. V. Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139-156). Amsterdam: Kluwer.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1). [Online: www.amstat.org/publications/jse/v3n1/konold.html]

Konold, C., Lohmeier, J., Pollatsek, A. Well, A. D., Falk, R., & Lipson, A. (1991). Novices' views on randomness. In R. G. Underhill (Ed.), *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 167-173). Blacksburg: Virginia Polytechnic Institute.

Laplace, P.-S. (1951). *A philosophical essay on probability* [English translation, original work published 1814: *Essai philosophique sur les probabilités*]. New York: Dover Publications.

Lecoutre, B. (2006). Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science*, *4*, 207-232.

Lecoutre, B., & Derzko, G. (2001). Asserting the smallness of effects in ANOVA. *Methods of Psychological Research*, *6*, 1-32. [Online: www.mpr-online.de]

Lecoutre, B., Lecoutre, M.-P., & Grouin, J.-M. (2001). A challenge for statistical instructors: Teaching Bayesian inference without discarding the "official" significance tests. *Bayesian Methods with Applications to Science, Policy and Official Statistics*, 301-310. Luxembourg: Office for Official Publications of the European Communities.

Lecoutre, B., Lecoutre, M.-P., & Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, *69*, 399-417.

Lecoutre, M.-P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, *23*, 557-568.

Loredo, T. J. (1990). From Laplace to Supernova SN 1987A: Bayesian inference in astrophysics. In P. F. Fougere (Ed.), *Maximum Entropy and Bayesian Methods* (pp. 81-142). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Moore, D. S. (1997). Bayes for Beginners? Some Pedagogical Questions. In S. Panchapakesan & N. Balakrishnan (Eds.), *Advances in statistical decision theory* (pp. 3-17). Boston: Birkhäuser.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330-357.

Piaget, J., & Inhelder, B. (1951). *La genèse de l'idée de hasard chez l'enfant.* [The origin of the idea of chance in children]. Paris: Presses Universitaires de France.

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximities data. *Psychometrika*, *47*, 3-24.

Robinson, D., & Foulds, L (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*, 131-147.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*, 319-345.

Savage, L. (1954). *The foundations of statistical inference*. New York: John Wiley & Sons.

Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics and learning* (pp. 465-494). New York: Macmillan.

Steinberg, H., & Von Harten, G. (1982). Learning from experience – Bayes' theorem: A model for stochastic learning. *Proceedings of the First International Conference of*

Teaching Statistics, Volume 2, (pp. 701-714). Sheffield, U.K.: Teaching Statistics Trust.

Thom, R. (1986). Preface of P.S. Laplace, *Essai philosophique sur les probabilités* [Text of the 5th edition, 1825]. Bourgois : Paris.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrate alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*, 371-386.

Vranas, P. (2001). Single-case probabilities and content-neutral norms: A reply to Gigerenzer. *Cognition*, *81*, 105-111.

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin*, *77*, 65-72.

MARIE-PAULE LECOUTRE
ERIS, Laboratoire Psy.Co, E.A. 1780
Université de Rouen, UFR Psychologie, Sociologie, Sciences de l'Education
76821 Mont-Saint-Aignan Cedex, France
http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris