Multiple perspectives on the concept of conditional probability

Manfred Borovcnik

Alps Adria University of Klagenfurt (Austria)

Perspectivas múltiples sobre el concepto de probabilidad condicional

Resumen

La probabilidad condicional es clave en la teoría subjetiva de la probabilidad; sin embargo juega un papel subsidiario en la concepción usual de la probabilidad, donde su contrapartida, es decir, la independencia, es de importancia básica. El artículo investiga estos conceptos desde varias perspectivas para arrojar luz sobre su carácter multifacético. Incluiremos las perspectivas matemática, filosófica y educativa. También se analizará la probabilidad condicional desde los ángulos de las ideas que compiten con ella y las estrategias de resolución. Para comprender la probabilidad condicional se requiere urgentemente una aproximación más amplia para superar los problemas bien conocidos de aprendizaje del concepto, que apenas parecen afectados por la enseñanza.

Palabras clave. Probabilidad condicional, probabilidad subjetiva, pensamiento probabilístico, falacias, paradojas en probabilidad.

Múltiplas perspetivas sobre o conceito de probabilidade condicionada

Resumo

A probabilidade condicionada é a chave para a teoria subjetivista de probabilidade; contudo, ela desempenha um papel subsidiário na conceção usual de probabilidade onde a sua contraparte, nomeadamente a independência, é de importância fundamental. No artigo investigam-se estes conceitos de várias perspetivas em ordem a lançar luz sobre o seu caráter multifacetado. Incluem-se as perspetivas matemática, filosófica e educacional. Além disso, inspecionamos a probabilidade condicionada a partir de ideias concorrentes e estratégias de resolução. Para a compreensão da probabilidade condicional é urgentemente necessária uma abordagem mais ampla para vencer os problemas bem conhecidos na sua aprendizagem, os quais parecem praticamente não serafetados pelo ensino.

Palavras-chave. Probabilidade condicional; probabilidade subjetiva; pensamento probabilístico; falácias; paradoxos em probabilidades.

Multiple perspectives on the concept of conditional probability

Abstract

Conditional probability is a key to the subjectivist theory of probability; however, it plays a subsidiary role in the usual conception of probability where its counterpart, namely independence is of

basic importance. The paper investigates these concepts from various perspectives in order to shed light on their multi-faceted character. We will include the mathematical, philosophical, and educational perspectives. Furthermore, we will inspect conditional probability from the corners of competing ideas and solving strategies. For the comprehension of conditional probability, a wider approach is urgently needed to overcome the well-known problems in learning the concepts, which seem nearly unaffected by teaching.

Keywords. Conditional probability, subjective probability, probabilistic thinking, fallacies, paradoxes in probability.

Multiples points de vue sur le concept de probabilité conditionnelle

Résumé

La probabilité conditionnelle est une des clés de la théorie subjectiviste des probabilités, mais elle joue un rôle secondaire dans la conception habituelle de probabilité où sa contrepartie, à savoir l'indépendance est d'une importance fondamentale. Cet article examine ces concepts à partir de perspectives différentes afin de faire la lumière sur leur caractère multiforme. Nous inclurons les perspectives mathématiques philosophiques et pédagogiques. En outre, nous allons inspecter la probabilité conditionnelle des angles d'idées concurrentes et les stratégies de résolution. Pour la compréhension de la probabilité conditionnelle, une approche plus large est urgente pour surmonter les problèmes bien connus dans l'apprentissage des concepts qui semblent presque pas affectée par l'enseignement.

Mots-clés. Probabilité conditionnelle; probabilité subjective; la pensée probabiliste; erreurs; paradoxes de la probabilité.

1. Introductory remarks

The usual answer to experienced difficulties in learning is to simplify the concepts and the choice with probability is the simulation method, which exploits the close relation between probabilities and relative frequencies.

While with probabilities the connection to relative frequencies might work, conditional probabilities remind us of the inherent dual character of probability as it adheres as a property to objects while at the same time it forms only a judgement in our creative mind: the object-subject duality defies simple reductions to one part. While the controversy in the foundations of probability and statistics has lost its fire and the philosophers let it unresolved – as in the foundations of mathematics after Gödel's theorems –, and in applications that concept is used, which is more helpful for the current problem, yet in teaching a biased approach is pursued, which reduces probability and conditional probability completely to some entity close to relative frequencies.

In this paper some ideas presented in an invited lecture at ICME 12 are elaborated (Borovcnik, 2012). We will start our analyses with an investigation from the mathematical and philosophical perspective (Section 2) and continue with a review of difficulties in teaching the concept and the current awareness of the problems (Section 3). The perspective of competing ideas (Section 4) will embed the rational mathematical approach into contending approaches such as causal readings of the situation, which are far more attractive to many people. To illustrate matters, we will use the Monty Hall problem. The perspective of solving strategies (Section 5) will categorize major types of strategies and explain cornerstones of their inherent view. In the conclusions (Section 6) we will discuss a programme to find a best teaching method

and will provide arguments for using a probabilistic approach instead of a simplifying one, as well as for developing a programme for further educational research.

2. A mathematical and philosophical perspective

What is randomness, or more specific, what is a random situation? The answer to this question changed by the development of formal concepts still is sensitive to the general cultural background and depends on the school of probability one adheres to. We will deal with crucial ingredients of both the objectivist and the subjectivist school, their key objects, and the theoretic justification of their basic ideas. Independence and conditional probability will accompany us as will Bayes' formula, which will serve as criterion for the adequateness of the like conception of probability. The philosophical perspective will lead to a dilemma between the demand of science to be objective and the demand of cognition to have a theory and procedures that extend and enhance thinking.

The multi-faceted character of probability as seen from a mathematical and philosophical perspective has been analyzed in several important works. Our background is formed by sources like Batanero, Henry, & Parzysz (2005), or Kapadia and Borovcnik (1991), and Borovcnik and Kapadia (2013a); for the controversy in the foundations of probability we learnt from Hacking (1975; 1990) and Stegmüller (1973). The bridge to history, especially to the history of physics is laid by von Plato (1994).

2.1. Objectivist setting of the theory of probability

Historically, probability was closely related to pro*v*ability: the more reasons one could find for a statement as compared to reasons against it, the more provable the statement was. Jacob Bernoulli (1713/1987) may be seen as tipping point in the emergence of probability as expressed by a number (between 0 and 1), which is *bound to an object*. Yet, a whole chapter of his *Ars conjectandi* is still devoted to the art of combining reasons for and against a statement (Shafer, 1996).

For the objectivists probability is separated from a person's judgements; it is a property inherent to an object like a physical property of it. The situations where such probabilities apply are indirectly determined by the axiomatic theory of Kolmogorov (1933/1957). Amongst others, the (weak) law of large numbers – already proven by Bernoulli – with the implicit assumption of an *independently* repeated binary experiment determines these situations. The relative frequency assumes a central function to measure probability replacing subjective weighing procedures as the corresponding random variable "converges" to the underlying probability.

This extends the scope of probability far beyond symmetric situations as in games of chance where weighing the evidence was essentially undisputed. It shifts also away the connotation from pro*v*ability to pro*b*ability as a property of the object as specified in the experiment. The repetition of random experiments is conceptualized by the concept of *independence*. While independence and dependence are defined by the same way via conditional probability, there is hardly an interest of its own in dependence within an objectivist framework or in investigating the interrelations between the unconditional, the joint, and the conditional probability.

For an event *E* with positive probability (i.e., $P(E) > 0$) the conditional probability of any (feasible) event in the sample space is defined as a probability measure, i.e., by

$$P_E(A) = P(A \mid E) = \frac{P(A \cap E)}{P(E)}.$$

This concept is justified completely by stating that the defined function P_E fulfils all the axioms of a probability measure and therefore *is* a probability (measure).

For an objectivist conception of probability, independence is key and crux

The special case of conditional and unconditional probabilities being equal, i.e., $P_E(A) = P(A)$ is called independence: *A* is called independent of *E*.

Independence enables the mathematician to develop a model for the repeated stochastic experiment which amounts to the basic paradigm of the objectivist position (*A* would then be an event described by the second and *E* by the first repetition of the experiment). Consistently, independence is the key assumption of central theorems as the law of large numbers or the central limit theorem (the convergence of standardized sums of random variables towards the normal distribution).

The rationale behind is mathematical consistency: Such a mathematical foundation helps to escape the trap of intuitions, which are abundant with probability. In terms of real situations where probability should be applied, independence is simply "explained" by a total lack of causal influence. Furthermore, a possibility for a validation of models is offered by statistical tests on experimental data to check whether independence applies or not in a specific situation. Critique against the concept and its usual justification within an objectivist framework is manifold.

First, the axiomatic approach of probability does not include statistical inference; this extension requires additional terms and criteria, which are disputed in the controversy about the foundations of statistics with a resulting dilemma of either accepting the irreparable gaps in the logical basis or admitting a subjectivist character of probability.

Second, statistical tests have to use such ambiguous criteria, which are multiplied with tests for independence as the main ingredients, which constitute tests (like the formulation of statistical alternatives and the control or optimizing for power) are not applicable.

Third, independence is an idealization which is void of an experimental test and is only justified by almost mystic arguments, or is set as an assumption (as opposed to a model): what happens if we assume the experiments were independent.

Fourth, as key concept, there should be some link to mental images or reference situations with independence – not only for the learner but also for an analysis within the philosophy of science. Any theory should justify specific interpretations of the key concepts and an analysis from this perspective has to check whether the theory can support the interpretation aimed for.

Fifth, for someone trained in the mathematics of objectivist probability, conditional probability – and independence – is a permanent source of confusion. The direct argument about lack of causal influence is vague and the validation argument by tests is theoretical and of little help to enhance the concept.

2.2. Subjectivist setting of the theory of probability

For a subjectivist, probability still is closer to provability: it is the personal expression of a degree of credibility of a statement, which forms the subjectivist counterpart of an event. A judgement of this credibility can be based on various types of information: i. Relative frequencies of relevant experiments in the past are one type of information, which is the sole input accepted by objectivists. ii. Information may also stem from expert knowledge: engineers, e.g., have failure data of technical units in an accelerated life test and re-transfer it to failure at normal stress; or, biometricians combine similar but differing empirical studies in a meta-analysis. iii. Such a degree can also be built up by personal expectations and therefore may be subjective.

The subjectivist position is also justified axiomatically. The axioms (de Finetti 1937; 1974) refer to rational behaviour and describe criteria for our preferences; the paradigmatic situation is betting on uncertain statements. Amongst others, preferences have to be transitive (if one prefers *A* to *B* and *B* to *C* then one has to prefer *A* to *C*) and one has to avoid bets that will cause a sure loss whatever is true. Calculating new from given probabilities is done in a mathematical manner; the difference to objectivist probabilists lies in the basis for probabilities, as beyond relative frequencies other ingredients play a role in the judgement of events.

Objectivists criticize this conception as the defined notion of probability *is* subjective and, furthermore, it is void of an empirical control by random experiments, which is a crucial requirement from the empiricist perspective on science. As reply to such a critique, there are three main arguments: i. For enough empirical data, subjectivists come to the same conclusions; ii. The theory by de Finetti has the same scientific status as the Kolmogorov theory as it is justified by the same method and it is also free of contradictions; and iii. Any judgement may be improved by further information and thus the probabilities get a more authoritative character on the way.

Any probability is conditional and updated by new evidence

While the basic character of probabilities is influenced by qualitative information, the subjectivist school of probability strives to increase the validity of – prior – probabilities used with various approaches: First, probabilities are updated by new evidence via Bayes' formula. Second, one should get enough data to minimize the influence of prior probabilities. Thirdly (e.g., Berger, 1993), a range of priors may be analyzed for its impact on the final decision and one speaks of robust priors if the influence is small for even larger changes of the prior probabilities.

As the evaluation of probabilities may be subjectively influenced, it becomes vital to integrate new knowledge – information – to form a better – updated – judgement of probability. This updating is ensured by Bayes' formula, which plays a central role: always unconditional (priors), joint, and conditional probabilities in both "directions" have to be regarded simultaneously. There is no wonder that Bayesians represent their key formula in a different way using odds as this fits better to the updating procedure. On the contrary, while Bayes' formula plays a key role in the subjectivist theory it seems to be a trivial theorem in Kolmogorov terms.

A further remark is due for the character of the theory and the meaning of concepts and statements: At least for less extreme representatives of the subjectivist school it is feasible to rewrite the subjectivist axiomatic theory in a form that parallels the Kolmogorov theory. However, the interpretation would remain different as would be

the approach to statistical inference. Unknown parameters of a distribution would have a prior distribution in the subjectivist setting and thus be a random variable; the relative likelihood of parameters with respect to an event E would be the value of evidence of E. The event E would be interpreted as statement, on which one could place a bet and which would thus have a probability.

Odds – relative weights instead of probabilities

Before Bernoulli it was usual to compare the cases in favour and the cases against an event: for a six with an ordinary die this yields a ratio of 1 to 5, or 1:5. This way of describing probabilities conforms also better to the wide-spread qualitative search for reasons for and against to judge the pro*v*ability of a statement in situations outside games of chance. With his golden theorem about the "convergence" of relative frequencies towards an unknown underlying probability, relative frequencies were upgraded and it became popular to denote probabilities by a number between 0 and 1; this calibration on a normalized scale was perceived as conceptual progress.

For many applications, however, a probability still may be better expressed by odds, i.e., the *relative* probabilities of an event to occur or fail to occur. For the six, the odds are 1:5 so that the fair stakes are easily derived as 5:1. While for additive probability calculations odds are clumsy, for joint and conditional probabilities odds allow for an easy calculus. We will restrict ourselves to the special case of two events H_i which form a partition of the fundamental set. We will refer to these events as hypotheses or simply states.

Bayes' formula with odds for calculating the posterior (conditional) odds – the number of reasons for H_1 as compared to against H_1 (i.e., for H_2) – relative to empirical evidence E reads as:

$$\frac{P(H_1 \mid E)}{P(H_2 \mid E)} \quad = \quad \frac{P(H_1)}{P(H_2)} \quad \times \quad \frac{P(E \mid H_1)}{P(E \mid H_2)}$$

$$\begin{array}{ccccc} New - posterior - odds & & old - prior - odds & & likelihood\ ratio \\ = & & \times & \\ (relative\ to\ empirical\ evidence) & & (without\ data) & & (power\ of\ data\ to\ separate\ states) \end{array}$$

Compare the formula to its standard notation

$$P(H_1 \mid E) \quad = \quad \frac{P(H_1) \cdot P(E \mid H_1)}{P(H_1) \cdot P(E \mid H_1) \ + \ P(H_2) \cdot P(E \mid H_2)}$$

and try to trace back or forward the influence of changes in the prior probabilities $P(H_i)$, the likelihoods $P(E \mid H_i)$ and the posterior probabilities $P(H_i \mid E)$: Neither the structural influence of single input factors, nor direction or the size of changes will be clear. On the contrary – with odds – the formula is well-adapted to the *multiplicative structure* of the situation: If prior odds or likelihood ratios double, posterior odds double; the higher the likelihood ratio $\dfrac{P(E \mid H_1)}{P(E \mid H_2)}$, the more probable is E under the state H_1 as compared to H_2. If a posterior probability is intuitively too small, this can be due to a very small prior, or due to weak evidence.

The consequence for the modelled situation is highly different depending on which of these factors is operative. For example, a diagnosing procedure in medicine is not per se ineffective if it leads to a posterior (after a positive result of the test) probability

to have the disease, which is too small. While the strength of evidence of a positive result may be very high, the final result may not be sufficient as the prevalence (the prior) is extraordinarily small as is with screening for HIV.

2.3. The philosophical dilemma

The subjectivist position is criticized for being subjective (!) while the objectivist side has gaps in the foundations of inferential statistics. The objectivist conception seems closer to the ideal of objectivity while the subjectivist theory embraces the inferential statistics with the help of Bayes' theorem. In the controversy on the foundations the decision was in favour of the objectivist conception in order to avoid any subjective connotation of probability (Stegmüller, 1973, or Hacking, 1975; 1990). In applications, statisticians nowadays use whatever is better for their problem and to a certain extent any application has subjective features.

3. An educational perspective on conditional probabilities

Conditional probability for objectivists is a special case of probability; it is justified as probability as it fulfils the axioms. Independence based on conditional probability is used for the formation of the theory: i. It is the key concept to define probability on infinite-dimensional spaces; ii. It is the key assumption to central theorems; iii. It is the basic ingredient of the concept of samples. For subjectivists, conditional probability is the basic concept per se: any probability is conditional to available information and is intrinsically tied to the idea of revising it by new evidence.

The Bayesian approach is much closer to how people think and can thus better explain the part of conditional probabilities. The clash between the philosophical schools overshadows the educational debate and blurs teaching efforts. We open this section by philosophical issues and how these are received within the mathematical community. Ensuing we note the exclusion of Bayesian ideas from curricula and round-off by some sources of difficulties to perceive conditional probability.

3.1. Awareness of philosophical debates within the mathematics community

The frequency interpretation of probability has an outstanding acceptance within *mathematical* statisticians, in spite of the famous and lively discussion by Berry (1997), Albert (1997), Moore (1997), and Witmer, Short, Lindley, Freedman, and Scheaffer (1997) in the *American Statistician*.

Consistently, teaching probability is dominated by the frequency idea. As it is not thus easy, amongst others the simulation approach has been developed and is now extensively used. While it enhances the relation between relative frequencies and probability, it enlarges the bias towards a frequency interpretation. Yet, the endeavour breaks down with conditional probabilities as many empirical studies show (e.g., Huerta, 2009; Díaz, & Batanero, 2009).

The reduction of probability to a one-sided notion tightly connected to frequencies cuts out only a small part of the multi-faceted concept of probability. As often, a simplified concept may be easier at first sight but by the end the learned concepts are misunderstood and abandoned later – not a good prospect for teaching.

It is remarkable how Bayesian ideas are excluded from curricula and even from didactic research, which may be seen from a discussion on urgent problems of teaching probability after a lecture at the ICMI/IASE Joint Study Conference in 2008 (see Borovcnik, 2011): Only one point of the elaborated agenda explicitly referred to subjective probability; conditional probability was not on the list even if it would pave the way to understand statistical inference. The analysis of Carranza and Kuzniak (2008) reveals the negative impact of neglecting Bayesian views in the curricula. However, it marks a solitary research piece amongst a community that strives for simplification.

3.2. Empirical evidence of difficulties

Conditional probability is known as a difficult concept. Context and used language of problems play a crucial role for the perception and reconstruction of the tasks. An overlap with causal thinking is often provoked if the context allows for such relations. Furthermore, a confusion of difficult mathematics with private thoughts may be traced with see-saw steps between these "worlds" as may be seen from a complexity analysis of items used in empirical research in Borovcnik and Bentz (1991).

Carranza and Kuzniak (2008) compose an educational dilemma from their analysis of curricula and textbooks: while the courses solely focus on the objectivist interpretation of probability as connected to frequencies, examples use extensively Bayes' formula and ask for a subjectivist concept. Studies by Díaz and Batanero (2009) reveal that mathematical knowledge does not really help; especially with problems of conditional probability there seems hardly a transfer of general features so that each problem and context has to be learned from anew. For us, this is a strong indication that education fails to reveal the inherent structure of such problems.

Language issues and problems of accepting mathematics

Language and idiosyncrasies including emotions are vital ingredients for the perception of situations with conditional probabilities. Empirical research suggests amongst others the following influence factors: slight variations in wording, the format (absolute numbers, proportions, percentages, or probabilities) of given and asked data, the context of the problems, and personal involvement both in context and solution.

All these factors influence how the students perceive a problem, what method they use to solve it, the way how they interpret the result and, finally, whether they *accept* the result. The matter of accepting is different from understanding and may block every attempt to teach or learn and is highly dependent on context: On the one hand, applications of conditional probability are nearly undisputed in medicine or law. Some rare cases have reached more public attention as the Simpson case (see Gigerenzer, 2002). On the other hand, problems like the Monty Hall (see below) arouse an unusual degree of emotions even by mathematicians involved – with an unjustified rejection of the official solution, which reveals a wide-spread lack of understanding the concept of conditional probability.

4. The perspective of competing ideas

Gut decisions are relatively successful (Gigerenzer, 2007). This seems surprising at first sight but in fact, there is no guarantee that a mathematical method would be more successful.

We will review three different types of ideas that are competing with a probabilistic approach: a direct prediction approach, captured by Konold (1989) in his outcome approach; the causal alternative to randomness and probability which seems very attractive for many people (Batanero, Henry, & Parzysz, 2005); a mixture of private conceptions and (partially understood) mathematical concepts. We will use extensively the Falk urn and the Monty Hall problem to illustrate our ideas. Finally, we will discuss paradoxical features of the tasks.

4.1. The prediction problem and Konold's outcome approach

An important criterion for a method for situations under uncertainty is that the method allows for an (exact) prediction of the *next* outcome. For this next outcome the probabilistic procedure and a private approach are compared by the degree to which they are comprehensible, by their potential to solve the prediction problem, and by the success afterwards.

A "wrong conception" might lead to success while a "correct model" might fail (see Borovcnik, 2011). After a failure of a private model a tendency for an excuse may be noted ("anyway it was only luck" etc.) and the next decision situation is perceived completely as *new*. Even if a series of such single-off decisions is re-evaluated after a while as something that belongs together, a bias in recall from memory will yield a useless measure for actual success.

There is yet the attractiveness of a strategy *before* the decision is made. The mathematical concepts fail to predict the next outcome (with certainty) and it is hard to understand what they really serve for. Another strategy might respond to the strong intuitive demand and predict the next outcome. According to Konold (1989), people tend to reformulate any probability statement into a direct (and absolutely certain) prediction of the outcome in question. He termed this idea "outcome approach".

4.2. The causal alternative to randomness

Assume the statement C is perceived causally and E is regarded as effect. Two ways of confusion are likely with the statement $P(C \mid E)$:

- The cause is thought to be independent of the effect: $P(C \mid E) = P(C)$.
  This reasoning is favoured in a time bound context when the cause precedes the effect and is dealt with below in the Falk urn example.

- Conditional probabilities are reversed as $P(C \mid E) = P(E \mid C)$.
  This phenomenon may be seen from an example on medical diagnosis by Tversky and Kahneman (1980).

Causes have to precede the effect: The Falk urn

Draw twice from an urn with two white and two black marbles (Figure 1). The drawn marble is not replaced. Two questions are posed (Falk and Konold, 1992):

1. You see the first marble to be white. What is the probability that the second marble will be white, too?

2. The first marble is temporarily hidden. Then, a second marble is drawn and proves to be white. What is the probability that the first marble is white?
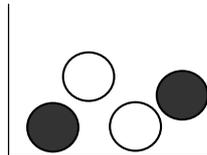

Figure 1. Falk urn with two white and two black marbles

While for the first problem the situation is directly represented by an urn yielding odds of 1:2 for white according to the current composition of the urn whence the probability is 1/3, for the second problem such an urn representation is missing. A variety of personal strategies are popular, especially an interference with causal perceptions (Borovcnik & Bentz, 2003, p. 49):

> "As for the first draw there were 2 white and 2 black marbles, the probability has to be 2/4 = 1/2 and the later events cannot influence the probability of the earlier."

This time-bound thinking is characteristic for the causal paradigm; the cause has to be first, the consequence later. Furthermore, variations to the Falk urn indicate that there seems to be a hierarchy between competing ideas: Private (often emotionally bound) reasons have the highest rank followed by logical reasons, then causal schemes, and at the lowest rank, probabilistic reasoning (Borovcnik & Peard, 1996).

Conditional probabilities simply reversed

For a disease D perceived as causing a positive medical test and a conditional probability of $P(+|D) = 0.99$ people tend to equate $P(D|+) = 0.99$ (Kahneman and Tversky, 1980). Of course, the *reversed* conditional probability depends on the prevalence of the disease; its value may be calculated by Bayes' formula. Such a prevalence differs hugely between subgroups.

Given that the calculus of conditional probabilities does correspond so badly to people's intuitions, competing non-probabilistic strategies or a mixture of probabilistic and other conceptions take over in Bayesian problems. Teaching strategies have to make it plausible that conditional probabilities have no link whatsoever to time and causes – they amount to a wider concept to integrate information of whatever kind into the probability evaluation of a statement.

4.3. A mixture of private and mathematical conceptions

Other private conceptions play a vital role and it is not easy to convince people of the lack of effectiveness of their strategies (Lysø, 2008). A mixture of partially understood mathematical concepts and intuitions can lead to re-formulations of the problems; furthermore, it can lead to a high confidence in a wrong perception of the situation. We will exemplify the situation by the well-known Monty Hall problem (see for example, Batanero, Fernandes, & Contreras, 2009; a valuable source for similar problems is Gardner, 1959, or 2006, and Székely, 1986).

Monty Hall problem

The main prize – a car – is hidden behind one of 3 curtains (doors) while behind the other curtains there is a goat (Figure 2). All curtains are closed at the beginning. The candidate chooses one. The moderator reveals a goat and offers to the candidate to change the initial choice.

Figure 2. Monty Hall – behind the curtains

As a good notation for a start, let A, B, C denote the doors and c_A signify the *state* that the car is behind door A. We assume that the candidate has chosen door A and the moderator has opened door C showing a goat. What is the probability that the candidate has selected that door with the car, i.e., what is the probability for c_A?

		states			
		c_A	c_B	c_C	
	A	0	0	0	0
moderator opens	**B**	1/6	0	1/3	1/2
	C	1/6	1/3	0	1/2
		1/3	1/3	1/3	1

Figure 3. Contingency table for Monty Hall if candidate chose door A

The solution may easily be read off from Figure 3. To fill in the entries in the table, we start with 1/3 in the margin at the bottom according to the random selection of a door. If the moderator has no preference for the other doors in case the candidate has chosen the car door, then doors B and C are opened with probability 1/2 so that the corresponding cells in the first column carry the entries of 1/6 each. The shaded cells are forbidden as the moderator never opens that door with the car. The solution can be read off the row related to the door the moderator opened; if C is opened, this yields $P(c_A \mid \text{"}C\text{"}) = \frac{1/6}{1/2} = 1/3$. Thus it would be better to change the initial choice.

What is paradoxical?

A fierce discussion followed the solution by vos Savant (n.d.). The degree of emotions is remarkable. Yet, there is a trap of equiprobability and the illusion to have logically exempted all information available and thus reached at the "correct" solution.

Emotions and responsibility. The more people are emotionally involved in the context, the more they are inclined to reconstruct the situation to *their* view. As may be seen from the many websites on this problem, people strongly tend to reject the idea – after a door is opened showing a goat – to change their initial choice.

Instead of analyzing the situation rationally – they start to think about *why* the option to change is offered to them; they want to know the deeper motives behind. They might also compare the situation afterwards: If they miss the car by their initial selection, then it is only bad luck; however, if they miss the car by their decision to change then it is their *responsibility*. Staying with the initial choice relieves the candidate from taking responsibility for personal decisions (Borovcnik, 2011, p. 77).

Illusion of using all information available. The information about the open door with the goat can be used to eliminate this door. Thus the fundamental space is reduced. Beforehand there were equal probabilities for all cases; it is tempting to regard the remaining doors as equally likely. This corroborates the private concern referred to above. At a *logical* basis – as people already reduced the space to two doors – this nourishes the *illusion* that all information has been utilized.

However, there is more evidence available to re-evaluate the remaining doors as the information differentiates between the remaining *states*. As that what has been observed has a considerably different probability under these states, the observation changes the equal probabilities.

A further source of confusion is that the presented solution refers to *states* rather than to where the car really is. If the candidate has chosen door *A*, c_A is often mistaken as "the candidate *has* the car" while it signifies that there is a probability of 1/3 that the candidate has chosen the car door. The solution proceeds only by hypothetical (and not real) states and the ensuing probabilities in the sense of "what happens if ...?"

5. The perspective of solving strategies

Solving strategies convey a lot about the concepts involved. There is also a close feedback and engrossing of understanding the concepts by following a particular strategy. Some strategies are more accessible to one type of learner, others are more successful to derive safely at a correct solution, others may enhance the concepts; some strategies are more promising at first sight but have their drawbacks later for conceptual progress.

Therefore, we will review a great variety of solving strategies and discuss their relative merits in the context of the Monty Hall problem. As above we denote the three different states (hypotheses) as c_A, c_B and c_C, the candidate has chosen door *A* while the moderator opened door *C* with a goat behind (statement g_C). The conditional probabilities of what we can see behind the opened door may be seen from Figure 5.

5.1. Naïve approaches and equally likely arguments

Here we will review some wrong strategies that are either personal or are characterized by a mixture of private and mathematical arguments.

Private reconstructions

Many people change the rules of the game insisting that only this perception of the rules is appropriate and the more emotional they react the higher is this tendency:

- The offer to change the initial choice seems unusual, the candidate might try to find a reason for it;

- The candidate does not like to take over any responsibility – the first selection is luck, a change of doors is an action and may be wrong;

- The candidate may have an emotional binding to the first choice – as if it were perceived as divine judgement and it were not advisable to challenge it.

Little room is left here for mathematical arguments. While the first strategy is surely more helpful in many similar situations, the other two remind us to universal strategies that embed probability in a general cultural background. In fact, chance was used since ancient times to outsource crucial decisions.

Equally likely arguments

On a logical level the opened door is eliminated. The two remaining doors "are" equally likely as they have been equiprobable from the beginning. This leads to 1/2 for both remaining doors and the conclusion that it is irrelevant whether one changes or remains with the first decision (Figure 4).

The approach is marked by a mixture of mathematical and private thinking. The logical exclusion is mixed with deep-seated equiprobability considerations. It gives them the illusion that they depleted all information. All these arguments might explain why people are highly convinced that they are right. Yet, the information is used only on a logical level to exclude the third door but it may serve further as an indication between the remaining doors.
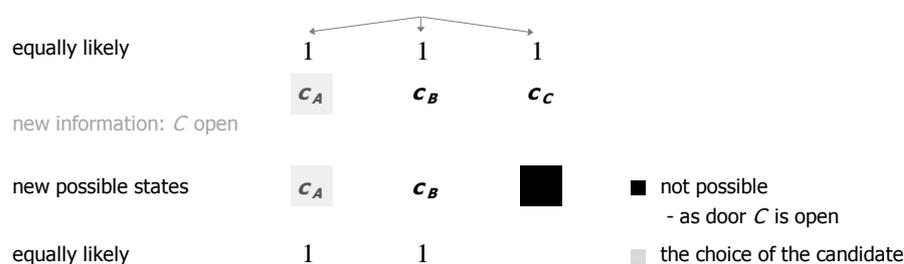

equally likely   1   1   1

c_A c_B c_C

new information: C open

new possible states c_A c_B ■ not possible - as door C is open

equally likely   1   1   ▨ the choice of the candidate

Figure 4. Equally likelihood arguments

5.2. Probabilistic strategies

We will review here mathematical strategies to solve the problem; one type centres around Bayes' formula, the other uses contingency tables for simplification.

Strategies based on Bayes' formula

The Monty Hall problem is a classical example of a Bayesian problem with prior probabilities for the three doors at the first stage and with empirical evidence at the second stage. One has only to fill in the corresponding (conditional) probabilities according to the context. We support this process by a tree diagram (Figure 5). The stages of the tree correspond to time: first the candidate chooses, then the moderator opens one door. As the moderator never opens the car door, there remains a free choice only if the candidate has hit the car by his choice A (i.e., c_A). From Figure 5, we read off:

$$\tfrac{1}{3} \cdot \tfrac{1}{2} = \tfrac{1}{6} \text{ for } c_A \wedge g_C \text{ and } \tfrac{1}{3} \cdot \tfrac{1}{2} + \tfrac{1}{3} \cdot 1 + \tfrac{1}{3} \cdot 0 = \tfrac{1}{2} \text{ for } g_C \text{ whence } P(c_A | g_C) = \tfrac{1}{6} / \tfrac{1}{2} = \tfrac{1}{3}.$$
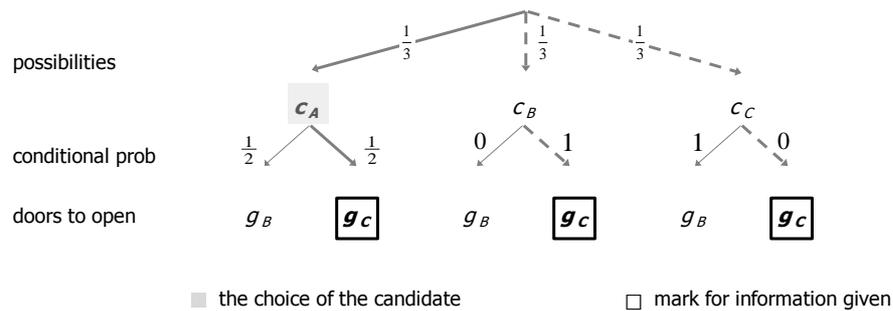

Figure 5. Bayes' formula in standard representation enhanced by tree diagrams

With odds the calculations proceed as follows: The prior odds of 1:1 for the possible states (hypotheses about whether the car is behind door A or not) and the likelihood ratio of 1/2:1 of the states under the evidence of door C open are transformed to posterior odds of 1/2:1 or 1:2 yielding a posterior probability of 1/3 for the candidate to have chosen the car door (see Figure 6).
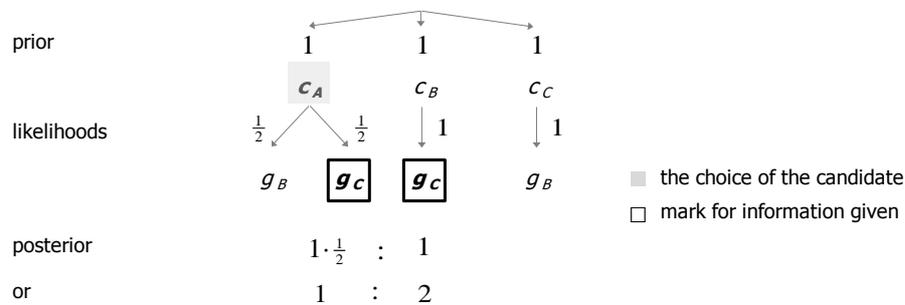

Figure 6. Bayes' formula in odds representation – enhanced by a tree diagram

This representation reveals more of the implicit structure of Bayesian problems and clarifies that the final – posterior – probability depends on the prior probability of the possible states (hypotheses) *and* the likelihood of the states relative to the evidence.

Strategies based on contingency tables

It is possible to use two-way tables, optionally with probabilities (as explained in Figure 3) or with absolute numbers to avoid the technicalities of Bayes' formula. In the following scenario expected numbers of a "statistical village of N people playing the game" are inserted according to the rules of the game (Figure 7):

- From $N = 300$ experiments one "expects" 100 for each of the states as the car is behind each door with the same probability.

- From the 100 cases with state c_B, i.e., the car is behind door B, all times the moderator has to open door C and show the goat there. From row C one may read off the solution as $50/150 = 1/3$.

		states			
		c_A	c_B	c_C	
	A	0	0	0	0
moderator opens	**B**	50	0	100	150
	C	50	100	0	150
		100	100	100	300

Figure 7. Contingency table for Monty Hall with absolute numbers if candidate chose door *A*

In the unit square method, the proportions of a contingency table are graphically represented. We identify absolute areas in the whole unit square and relative area in smaller figures (see Figure 8):

- Vertically, the initial states allocate stripes of 1/3 proportional to their probability (left hand figure);

- In a second step, and horizontally, conditional probabilities of the evidence (goat behind door C) split the stripes in two parts.

The total shaded area stands for the evidence. The conditional probability of one state is represented by the corresponding shaded area within the stripe related to that state relative to the whole shaded area. Here, the shaded area has three equal parts, one of which corresponds to c_A. Thus, the conditional probability to have the car is 1/3.
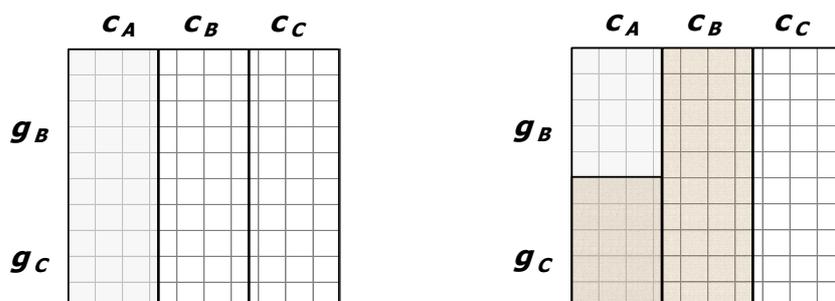

Figure 8. Unit square method – the two hierarchic steps to split the figure if the candidate chose door *A*

5.3. Adequate non-probabilistic strategies

Here, two correct strategies are reviewed, which solve the problem without explicit use of probabilities.

Tree diagrams with absolute numbers

As with the contingency tables in the last example, we can use tree diagrams with absolute numbers as a graphical representation of possible cases. This strategy may be pursued without using probabilities. At the first level the split is between the three states, at the second the split is according to the rules (see Figure 9). If the car is behind door *A* (the candidate's choice) then there is a random choice between doors *B* and *C*.
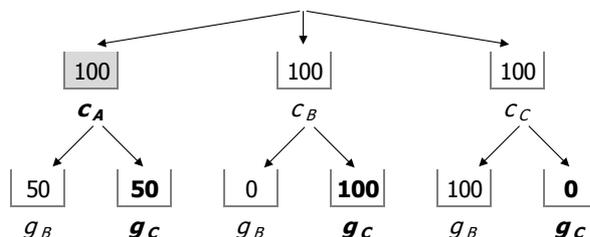

Figure 9. Tree diagrams with absolute numbers if candidate chose door *A*

In the containers corresponding to the opened door *C* there are all in all 150; from these 50 come from the candidate's choice whence the solution is 50/150 = 1/3. With the idea of random selection this fraction may be interpreted as probability.

Graphical representation of the possible cases

By a symmetry argument, the possible actual constellations of the choice of the candidate may be reduced to the three equally likely cases shown in Figure 10. For the attribution of "goat, car, goat" to the doors *A* to *C*, the candidate chooses without preferences so that the three lines of the figure represent equally likely cases. The trick of using a schematic spatial drawing allows to look behind the scenes (doors). Only in the second line (candidate has hit the car), the moderator has a free choice to open a door. Now in two of three (equal) cases, changing doors wins while remaining with the first decision only one of three cases is favourable whence the probability to have the car is 1/3 (Gigerenzer, 2002).
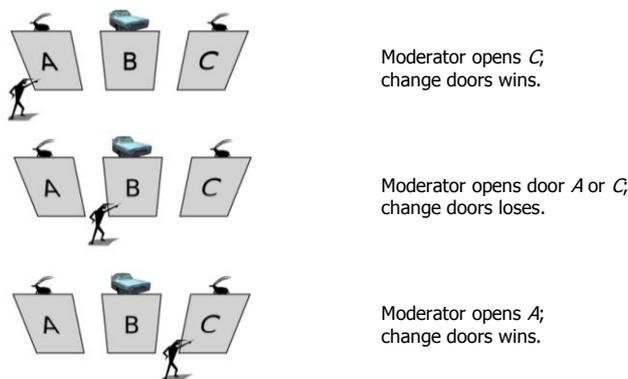

Figure 10. Solution by a schematic plot of all cases

6. Conclusions

We start our conclusions with a debate on the old didactic vision to find a best teaching strategy. We advocate for probabilistic approaches towards conditional probabilities, at least from intermediary steps onwards. An agenda for further educational research will touch also more general points beyond conditional probabilities.

6.1. Search for a best teaching method

Gigerenzer designed a programme to investigate, which embodiment is easier to understand and with co-workers he performed many empirical studies (Hoffrage, Gigerenzer, Krauss, & Martignon, 2002; Sedelmeier & Gigerenzer, 2001). The tree diagram with absolute numbers is found most effective followed by tables with absolute numbers[1]. Both methods avoid the calculus of Bayes' formula; they reformulate a probability task in terms of frequencies, solve it by numbers and re-interpret the solution legitimately as probability. They do not even explicitly use conditional probabilities as the reference set is clear from the context.

Nonetheless, the approach misses:

- To clarify the structure of the situation (e.g., why a conditional probability is so small);

- To offer an iterative application as is required often in Bayesian problems;

- To investigate the influence of variations of input parameters on the result (often only unreliable estimates are available);

- To provide possibilities to check for a change in the assumptions (e.g., a change of prior probabilities).

Of course, the basic correct solution is a strong argument for Gigerenzer's natural frequencies and the tree approach. Yet, more investigations and *deliberations* have to be done how to shape the curriculum as progress in stability to derive a numerically correct solution is only *one* criterion for success.

The value of the various methods depends highly on the aim of teaching – conceptual or numerical progress – and even on the person to teach. Tree diagrams have a distinct order and distort the symmetry between the two characteristics since order of levels are not open to reverse. Two-way tables (with absolute numbers) allow for interchanging rows and columns which corresponds to the symmetric character of conditional probabilities. Also, either causal or diagnostic, both "directions" are conceptualized by conditional probabilities. Another feature of Bayesian reasoning is its iterated application, which is better captured by the odds representation.

6.2. Probabilistic strategies to teach conditional probabilities

At higher steps some fluency with – conditional – probability rules gets required and will be missing if the method is avoided earlier. Probabilities and frequencies are

---

[1] In fact, a discretized form of the unit square was found even better but required technological support.

on a complete different conceptual level. The absolute numbers represent an *expected* value; as a natural number it gives the illusion of exactness, which conceals that the source often is an *estimate* of the probability with a big margin of statistical error.

For small probabilities this problem is enlarged as the quality of data is unreliable: First, the assumptions of a random sample are rarely met. Second, the variation in the sample (if it *were* random) is big. For a probability of $p = 10^{-4}$, among a sample of 10000 there is a 36.8% chance to get *no one* with the property with an estimate of $\hat{p} = 0$, a further 36.8% for an estimate of $\hat{p} = 0.0001$ (exact), 18.4% for an estimate of $\hat{p} = 0.0002$ (doubles the value), 6.1% for $\hat{p} = 0.0003$ (triples it) and 1.9% for an estimate more than four times the underlying value.

Furthermore, numbers attract the character of facts while in many cases probabilities reflect more or less the status of – controversial – hypotheses. Thus, for teaching it is vital to deliberate and re-consider the aim. Simplifying helps the majority to success for the short term but may bias the concept.

Several strategies may enhance learning of conditional probabilities and the Bayesian formalism.

- Implicit or hidden lotteries (Freudenthal, 1973; Bentz, 1983). The lottery on the choice of the doors (which door hides the car) is symmetric but the implicit lottery on the choice of the moderator violates symmetry.

- The favour concept (Falk & Bar-Hillel, 1983; Borovcnik, 1987). This concept discusses qualitative relations between circumstantial evidence and states, which are essential for inductive reasoning. The actually opened door does not favour (nor disfavour) to have the car whence its probability remains unaltered with 1/3. However, the open door is circumstantial evidence for the still closed door to host the car.

- Re-presentation of formalism (Borovcnik & Peard 1996). Bayes' formula with odds fits better to the idea of revising probabilities and connects objectivist and subjectivist conceptions more closely. Given evidence has a strong power to separate between the states if it has a high (conditional) probability under one of the states and low probabilities under the other states.

If the perspective of teaching conditional probability is to improve probabilistic thinking (Borovcnik 2006), teaching strategies should also react to Carranza and Kuzniak's (2008) fundamental critique that subjectivist connotations have to become an integral part of probability teaching.

6.3. A programme for further educational research

Here, some points of an agenda will be depicted, which might enrich the discussion about teaching conditional probability.

Competing ideas and conceptions – clarify or dismiss?

It is inconsistent to ignore other approaches or dismiss them simply as irrational. Yet often in teaching mathematics, intuitions or private conceptions are regarded as a hindrance to direct access to the concepts. With probability, the negative effect of such an attitude is even worse as it is well-known that learners resort to their old – nearly

unchanged – conceptions in out-of-school contexts.

To use an interventional interview in early phases of teaching probability may provoke the learners to openly discuss about their conceptions. This is an opportunity to confront them with variations of the context, with other problems or ideas to show the limitations of their thought. Such a strategy was suggested by Borovcnik and Peard (1996) and is used by Lysø (2008).

Modelling ideas – clarify or complicate matters?

With modelling, more mathematics usually is needed; rote learning never really helps but clearly fails with modelling. Yet, modelling may clarify the concepts better: First, the purpose of probabilistic methods seems widely misunderstood. It is not possible to predict an outcome with certainty as people would expect intuitively. However, according to Konold (1989) any probability statement is reformulated to enable such a prediction.

Second, rarely, tasks are posed where students experience how modelling improves a situation for a decision according to a specified criterion; most tasks ask to calculate some probabilities and interpret them in a – nearly arbitrary – context. Borovcnik and Kapadia (2011), e.g., illustrate the modelling potential of probability.

Third, probability is a modelling entity; it is *not* a property of real objects. The probability for a six with dices is often misread as 1/6. It is not 1/6! We get 1/6 only if we use a *model* of equiprobability. We could use another model which might better reflect the asymmetric figure of the holes in the die.

Some basic conceptual confusion

First, a confusion between hypotheses and data: For coin tossing, e.g., the model used comprises independent trials with equal probabilities. Based on this model, data with four heads in a series, has no effect on the next outcome. If the model is strong, then data induces no re-evaluation of it. Yet, in applications, awkward data offers a challenge to re-consider the model used: students react and use the pattern of the past to predict the next outcome. The crucial question from a modelling perspective is, when does the data indicate that the model should be altered?

Second, the confusion of ideally repeated situations and one-off decisions: Most situations – even in games of chance – are either one-off or perceived as such by people. By the sole focus on the frequentist interpretation (and on simulation) in teaching, misleading conceptions are shaped: either students are confused or return to private strategies in their decision making. For a discussion of the problem of taking out an insurance policy under this perspective, see Borovcnik (2006).

Third, the confusion of subjective and subjectivist: Subjectivist probabilities are dismissed as arbitrary, although usually subjectivist probabilities are based on expert knowledge. As a consequence, subjectivist connotations are ignored in teaching but may help to overcome many paradoxes especially those linked to conditional probabilities. It seems a prime goal to develop a blended approach of subjectivist and objectivist elements. Barnett (1973) has undertaken a comparative approach as a potential to understand both and appreciate their relative merits but only recently, Vancsó (2009) designed a parallel course developing classical and Bayesian statistics side by side following a similar intention.

Fourth, the merging of probability and impact: To make probability more

challenging, more recently suggestions have been made to start probability teaching by risk evaluations (Kent, Pratt, Levinson, Yogui, & Kapadia, 2010). With their experiments, Kahneman and Tversky (1979) have shown that people react highly sensibly to changes in the setting regarding to small probabilities if they are related to high impacts. Their "irrational" choices in such settings may be taken as an indication that for learning, the internal probability scale has to be calibrated first before we can proceed to integrate impact into teaching. The dilemma obviously lies in the fact that only in abstract settings a perception of impact – attributed to the situation by the persons within their own reconstruction – may be avoided. Thus, the task of calibrating one's probability scale meets the superposition of impact to some degree.

Orientation of research towards empirical studies vs ideas

In our data-driven world any knowledge has to be evidence-based. Consistently, current educational research is strongly oriented towards empirical investigations. Methodologically sloppy steps in planning and performing such "educational experiments" and the confounding factor of an experiment with – groups of – human beings reduce the value of the final conclusions considerably.

Huerta (2009), e.g., has convincingly shown that empirical research on conditional probability is limited in scope as the types of problems in studies exhaust only a minority of the huge variety of problems. Furthermore, pedagogic considerations are extended into the discipline (of mathematics) leaving their scope and potential behind them. To re-establish a balance between the subject and its pedagogy is an urgent point of an agenda to shape teachers' knowledge. The value of a rational discourse on ideas behind the concepts from a perspective of integrating philosophy of science, mathematics, history, cognitive psychology, etc. has to be re-considered in comparison to the dominating evidence-based orientation of research.

Clarify the conditional character of basic probabilities in statistical inference

Many researchers as Borovcnik (2011) insist that a sound understanding of conditional probability is vital for probability education, "Clarification and demystification of conditional probability from concurrent causal ideas is a pre-requisite for understanding concepts of inferential statistics like type I and II errors, or *p* values" (Borovcnik, 2011, p. 80).

The usual approach in teaching statistical inference by simulation studies obfuscates their character as *conditional* probabilities: First, the effect of causal connotations is ignored. Second, the frequentist meaning of probabilities is valid only within an artificial context that misfits the test situation; the discrepancy between the real situation and the model actually used is amplified in this *iid* scenario[2] to all other repetitions of the selected model. Third, the simulated numbers camouflage their tight connection to the assumptions underlying the simulation study, which blurs the genuine *conditional* character of the simulated probabilities.

The ongoing puzzles[3] linked to conditional probability (see e.g., Vancsó, 2009) are a sign that the restriction of probability to (the limit of) relative frequencies fails to

---

[2] *Iid:* independently and identically distributed is the paradigmatic assumption to random experiments.

[3] The reader might find more about the value of paradoxes to build probabilistic concepts in Borovcnik and Kapadia (2013b).

build stable conceptions in the learners. Despite the current trend towards data handling in international curricula, probability - and especially conditional probability - is still relevant as may be seen from Jones (2005). More deliberations are needed how to integrate subjectivist ideas into the probabilistic curricula in order to overcome the well-known shortcomings of current approaches.

References

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, *51*(3), 247-253.

Batanero, C., Fernandes, J. A., & Contreras, J. M. (2009). Un análisis semiótico del problema de Monty Hall e implicaciones didácticas (A semiotic analysis of the Monty Hall problem and didactic implications). *Suma*, 62, 11-18.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 15-37). New York: Springer.

Barnett, V. (1973). *Comparative statistical inference*. New York: Wiley.

Bentz, H.-J. (Ed.) (1983). Probleme im Umgang mit dem Zufall (Problems in dealing with randomness.). *Der Mathematik-Unterricht*, *29*(1).

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Bernoulli, J. (1713/1987). *Ars conjectandi*. Basel: Impensis Thurnisiorum, Fratrun; translation of 4th part by N. Meunier. Rouen: Institut de Recherche sur l'Enseignement Mathematique (Original work published in 1713).

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*(3), 241-246.

Borovcnik, M. (1987). Revising probabilities according to new information – A fundamental stochastic intuition. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 298-302). Victoria: University of Victoria. Online: www.stat.auckland.ac.nz/~iase/publications/icots2/Borovcnik.pdf.

Borovcnik, M. (2006). Probabilistic and statistical thinking. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education* (pp. 484-506). Barcelona: European Society for Research in Mathematics Education. Online: ermeweb.free.fr/CERME4/.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education: A joint ICMI/IASE Study* (pp. 71-83). New York: Springer.

Borovcnik, M. (2012). Conditional probability – a review of mathematical, philosophical, and educational perspectives. Paper presented at the Topic Study Group 11 "Teaching and Learning Probability" at ICME 12, Seoul. Online: www.icme12.org/sub/tsg/tsgload.asp?tsgNo=11.

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

Borovcnik, M., & Bentz, H.-J. (1990/2003). *Intuitive Vorstellungen von Wahrscheinlichkeitskonzepten: Fragebögen und Tiefeninterviews* (Intuitive conceptions of probabilistic concepts: Questionnaire and in-depth interviews). Technical Reports. Klagenfurt University.

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education* (pp. 239-288). Dordrecht: Kluwer.

Borovcnik M., & Kapadia, R. (2011). Modelling in probability and statistics – Key ideas and innovative examples. In J. Maaß, & J. O'Donoghue (Eds.), *Real-world problems for secondary school students – Case studies* (pp. 1-44). Rotterdam: Sense Publishers.

Borovcnik, M., & Kapadia, R. (2013a). A historical and philosophical perspective on probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education*. New York: Springer – to appear.

Borovcnik, M., & Kapadia, R. (2013b). From puzzles and paradoxes to concepts in probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education.* New York: Springer – to appear.

Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in french education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Proceedings of the Joint ICMI/IASE Study: Teaching Statistics in School Mathematics*. Granada: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annals of the Institute Henri Poincaré*, *7*, 1-68.

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

Díaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, *4*(3), 131-162. Online: www.iejme.com/032009/main.htm.

Falk, R., & Bar-Hillel, M. (1983). Probabilistic dependence between events. *Two-Year College Mathematics Journal*, *14,* 240–247.

Falk, R., & Konold, C. (1992). The psychology of learning probability. In F. Sheldon & G. Sheldon (Eds.), *Statistics for the Twenty-First Century*, MAA Notes 26 (pp. 151-164). Washington D. C.: The Mathematical Association of America.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel.

Gardner, M. (1959). *The Scientific American book of mathematical puzzles & diversions*. New York: Simon & Schuster.

Gardner, M. (2006). *Aha! A two volume collection: aha! Gotcha aha! Insight*. Washington, DC: Mathematical Association of America.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking.

Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition 84*, 343–352.

Huerta, P. (2009). On conditional probability problem solving research – Structures and contexts. *International Electronic Journal in Mathematics Education*, *4*(3), 163-194. Online: http://www.iejme.com/032009/main.htm.

Jones, G. A. (2005). *Exploring probability in schools. Challenges for teaching and learning. Mathematics Education Library, Vol 40*. New York: Springer.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *XLVII,* 263-291.

Kapadia, R., & Borovcnik M. (Eds.) (1991). *Chance encounters. Mathematics Education Library volume 12*. Dordrecht: Kluwer Academic Publishers.

Kent, P., Pratt, D., Levinson, R., Yogui, C., & Kapadia, R. (2010). Teaching uncertainty and risk in mathematics and science. In C. Reading (Ed.), *Proceedings of the Eight International Conference on Teaching Statistics*. Voorburg: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications.php ?show=icots8.

Kolmogorov, A.N (1933/1956). *Foundations of the theory of probability*. London: Chelsea.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59-98.

Lysø, K. O. (2008). Strengths and limitations of informal conceptions in introductory probability courses for future lower secondary teachers. In M. Borovcnik, D. Pratt, Y. Wu, & C. Batanero (Eds.), *Proceedings of TSG 13 on 'Research and development in the teaching and learning of probability'*. Monterrey: ICME 11. Online: www.stat.auckland.ac.nz/~iase/publications.php?show=icme11.

Moore D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254-261.

Sedelmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General 130,* 380-400.

Shafer, G. (1996). The significance of Jacob Bernoulli's Ars Conjectandi for the philosophy of probability today. *Journal of Econometrics*, *75*, 15-32. Online: http://www.sciencedirect.com/science/journal/03044076/75/1.

Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, vol.4, first part: Personelle Wahrscheinlichkeit und Rationale Entscheidung, second part: Personelle und statistische Wahrscheinlichkeit. Berlin-New York: Springer.

Székely, G. J. (1986). *Paradoxes in probability and mathematical statistics*. Dordrecht/Boston: D. Reidel.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgment under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, N.J.: L. Erlbaum.

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, 4(3), 291-322. Online: www.iejme.com/032009/main.htm.

von Plato, J. (1994). *Creating modern probability: Its mathematics, physics and philosophy in historical perspective*. Cambridge: Cambridge University Press.

vos Savant, M. (n.d.). *Game show problem*. Online: www.marilynvossavant.com/articles/gameshow.html.

Witmer, J., Short, T. H., Lindley, D. V. Freedman, D. A., & Scheaffer, R.L. (1997). Teacher's corner. Discussion of papers by D. Berry, J. Albert, and D. Moore, with replies from the authors. *The American Statistician*, *51*(3), 262-274.

Author

Manfred Borovcnik. Department of Statistics, University of Klagenfurt (Austria).

manfred.borovcnik@uni-klu.ac.at