6-2015

Risk and Decision Making: The “Logic” of Probability

Manfred Borovcnik

Follow this and additional works at: <https://scholarworks.umt.edu/tme>

Part of the [Mathematics Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Borovcnik, Manfred (2015) "Risk and Decision Making: The “Logic” of Probability," *The Mathematics Enthusiast*: Vol. 12 : No. 1 , Article 14.

DOI: <https://doi.org/10.54870/1551-3440.1339>

Available at: <https://scholarworks.umt.edu/tme/vol12/iss1/14>

This Article is brought to you for free and open access by ScholarWorks at University of Montana. It has been accepted for inclusion in The Mathematics Enthusiast by an authorized editor of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

Risk and Decision Making: The “Logic” of Probability

Manfred Borovcnik

Alpen-Adria University, Klagenfurt, Austria

Abstract: Risk is a hot topic. There is an international trend to use examples of risk or the concept of risk in the early teaching of probability. It enriches the problems, it widens the contexts, and it motivates the students to learn probability. This paper illustrates the notion of risk as a multi-faceted concept. The diverse perceptions start with language where risk is used in very different ways. The overlap of risk and hazard is not restricted to the technical context of safety and reliability; Knight’s seminal work on risk and uncertainty has its definite impact on today’s perception of the notions. The endeavour to re-interpret issues of statistical inference by risk – the risk of type I and II errors – or the concept of the weighted impact of decisions (in decision theory and in Bayesian framework) can clarify what risk means within mathematics but – as the whole machinery of statistical inference is difficult to understand – it may have little consequence on how the majority of people act and understand the notion of risk. Kahneman and Tversky show that the perception of risk is influenced by psychological factors and assert that people are risk averse in winning situations while they are risk seeking in losing situations. The perception of risk is dominated by the impact (loss or win) so that even a thorough judgement of the underlying probabilities is biased. If risk is shared between several stakeholders, they all have to use their own ingredients for their model of the same situation and follow their own logic. This leads to non-unique answers, which is unusual in mathematics. Methods of simplifying problems, the way to find a solution, and understand the underlying concepts more easily may induce a shift from a refined perception of the (hypothetical) models involved towards factual knowledge. The article aims to clarify these issues, which influence the ways to conceptualize, perceive, and teach the notion of risk.

Keywords: risk, uncertainty, risk perception, decision making, statistical errors, Bayesian risk, minimax principle.

Introduction

The reader might think that risk has a well-defined meaning and can be taught straightforwardly. The following discussion illustrates how risk can enrich probability teaching but reminds us that deliberate learning paths have to be designed so that teaching can enhance the ideas of risk in the learners. Modelling requires a different “logic” adapted to the problem being modelled and its context, which seems surprising for many as they do not perceive such diversity in mathematics and are also surprised that probability does not follow logical rules such as transitivity. In an earlier paper on risk in health we found it necessary to define risk right at the beginning of the essay:

“By risk we understand a situation with inherent uncertainty about the (future) outcomes, which is related to impact (cost, damage, or benefit). Sometimes expected value is used for comparing several decisions, which are ‘at stake’. Risk is used heterogeneously, some refer only to the probability inherent to one adverse outcome without regarding its impact, and others refer only to the adverse outcome. A decision between several choices of action might involve one person, or a decision might be ‘shared’ between two or more stakeholders, e.g., a patient and a doctor who have to find a decision about the next steps. [...] Risk involves two components and both

are prone to subjective interpretations: the judgment of impact is different for different individuals and is even more distinct for a person or an institution with a role different from the patient.” (Borovcnik, & Kapadia, 2011a, p.1)

However, one reviewer criticized our usage of the term “risk” and stated that by risk only the adverse future outcome can be meant. This would also be reflected by sentences like “The operation bears the risk of ...”, “I do not climb in the mountains as the risk of an accident is ...”, “No risk – no fun”. This shows that there is no unique concept of risk and we became aware that there are other feasible ways to define and think about risk instead of a precise definition.

Therefore, we start with a discussion of various meanings of risk and also refer to concepts linked to risk and describe situations that involve risk. In the second section we identify statistical notions of risk, which include type I and type II errors of statistical tests and Bayesian risk. The third section is devoted to psychological aspects of risk and perception of risk. We continue with three paradigmatic examples of risk that develop genuine features of risk and enrich the notion of probability by their context. In a further section, elementary approaches to probability are intended to make sense of risk involved in specific situations. A special view on the stakeholders that are possibly brought together in decision making corroborates our view that there are many ways to model and describe risk, and they all yield feasible descriptions of the situations involved but may differ in character and in what is regarded as the best decision. Implications for teaching complete the multifaceted character of risk that we project by our considerations.

Meanings of Risk

In this section, we deal with various ways to define risk, which is unusual from a mathematical perspective, where definitions are usually agreed. Furthermore, the meaning of risk – regardless of the specific definition – is also influenced by the character of probability information that is used for determining risk. We also go into details about relations between risk and hazard, between risk and uncertainty, and risk and utility. This section concludes with an analysis of situations that involve risk. The constituent parts of such situations (like who takes a risk) influence the interpretation of the input information as well as the meaning of the calculated risk.

Attempts to Define Risk

Apart from the vague use of “risk” in everyday language, risk is perceived and defined also in the scientific context. The common element between science and everyday seems to be an unwanted event that may or may not occur, often related to a low probability and severe consequences. Hansson (2007) discriminates five definitions of risk:

- (1) “an *unwanted event* which may or may not occur”;
- (2) “the *cause* of an unwanted event which may or may not occur”;
- (3) “the *probability* of an unwanted event which may or may not occur”;
- (4) “the statistical *expectation value* of an unwanted event which may or may not occur”;
- (5) “the fact that a decision is made under conditions of *known* [rather than unknown] *probabilities*”.

An early definition of “risk” in the sense of (5) has been attempted by Knight (1921) who made the term very popular; his use has its roots in economic settings (see below). Generally, the use of

probability, which is also controversial, is essential for the perception and interpretation of risk. Indeed, we refer to the three main classifications of probability in Borovcnik and Kapadia (2014), which are useful to clarify discussions on probability-related issues. We argue that all three approaches should be introduced in teaching probability. We will use this helpful classification throughout the paper and hope that others involved in teaching probability will use this terminology.

APT: A priori theory – probability is identified by equal probabilities in a finite probability space (Laplace theory).

FQT: Frequentist theory – probability is linked (determined, defined) by the limit of relative frequencies.

SJT: Subjectivist theory – probability is the personal degree of belief of a person in an uncertain statement.

There is an axiomatic justification for each approach and its inherent interpretation. The three approaches have led to fierce controversies and disputes historically. This is partly because there are restrictions for each of them. There are no situations with equal probabilities, there is no limit of relative frequencies in the finite series of experiments in the real world, and the “elicitation of probabilities” is too subjective for many people. Besides and along all these approaches, probability is also used as a scenario term to investigate a real problem on the basis “what happens if ...” (see Borovcnik, 2006; 2011). An example of the use of the scenario idea is the following:

How many redundant technical units (built in a system in parallel so that all have to fail for the system to fail) are required to attain an acceptable risk (probability) of failure of the system *if* the reliability (probability of functioning) of single units is 0.95 and the failures are independent? Mathematically, one can calculate the system’s reliability as $1-0.05^n$ and compare the increase in reliability to the increase in cost and decide about a reasonable balancing point. The values calculated allow for a transparent decision but are not to be taken literally as approximating any realistic probability (neither as property of single units nor of the system).

The meaning (4) about risk is the one preferred in technical environments dealing with the reliability of systems. It seems more objective and thus widely communicable or compulsive. However, its relevance still depends on the actual probabilities involved. There might be limited documentation about failures of units, there might be only qualitative statements of engineers (from stress tests under conditions of highly elevated stress levels, which have an effect on shorter life time and earlier failures), and there might be only general considerations of redundancy as in the scenario above. In cases when there is a very low probability of failure of a unit (which is often true), there are no failures to observe in real life under normal conditions so that any probability is a scenario figure.

Though the discussion of risk has been initiated by Knight in an economic setting, the technical usage of risk has been made popular by early studies on nuclear reactor safety (Rasmussen et al., 1975). It is important to regard the character of probabilities involved in risks that are communicated as well as the precision of knowledge on the probabilities used to calculate the risk. To speak in terms of risk about probabilistic situations is just a twist in language. Similar problems have been dealt with since the early beginnings of probability. And they have led to unsolvable paradoxes like the St Petersburg paradox, in which the fair stakes of a bet are derived as an infinite amount (see Borovcnik, & Kapadia, 2014). This also prompted considerations of the utility of money for the first time, by which the expected value of the game turns to the risk in the sense of (4). The change of terms, however, reflects also a change of applications of probability, especially to the economy (where risk is an inherent component of actions)

and reliability and safety considerations in technology (where risk is also shared between stakeholders of unequal access to influencing the decision).

Risk and Hazard

Various dictionaries (like Merriam-Webster) also link the adverse event with the potential underlying cause and subsume it under the heading of “risk”, such as the possibility that something bad or unpleasant (such as an injury or a loss) will happen, someone or something that may cause something bad or unpleasant to happen, or a person or thing that someone judges to be a good or bad choice for insurance, or a loan.

There are key differences in the five definitions in the previous section. In (1) there is an unwanted event; in (3) it is the probability of an adverse event; a summary figure for risk occurs in (4); in (2), risk is *something that may cause an adverse event*. There is a subtle difference of wording in (1) and (2) with the latter focusing on the factors behind, which might cause the harm. Such factors are called *hazards*. However, usage of the notion hazard is diverse outside technical contexts. Often dictionaries do not give specific definitions of it or combine it with the term “risk”. For example, one dictionary defines hazard as “a danger or risk” or refers risk back to hazard which helps explain why many people use all the terms interchangeably and contribute to the confusion. Relating to workplaces, CCOHS (n.d.) states: “A hazard is any source of potential damage, harm or adverse health effects on something or someone under certain conditions at work.” Examples of hazards at the workplace are materials like asbestos that may cause mesothelioma, or substances like benzene that may cause leukaemia.

In this setting, hazard becomes what earlier has been subsumed in (2) as risk, and risk becomes the chance or probability that a person will be harmed if exposed to the underlying hazard; this is (3), not (4). Risk assessment becomes the process of identifying hazards and evaluating the risk associated to the specific hazard; risk management is then any measurement to eliminate or control hazards at an acceptable level.

The problem with hazards is that they are linked via probabilities or chance to the adverse events. If the effect is acute (with immediate consequences) the link back to the hazard is obvious. However, more often the effect is chronic and delayed so that a link is hard to establish and it is difficult to provide evidence for. Potential hazards are statistically identified by correlations or associations (ex post exposure rates) and rarely can a causal link be established. That makes it very difficult to convince people about (potential) hazards.

When we study the relationship between smoking and the occurrence of lung cancer, we notice that the percentage of people with lung cancer is greater in those who have smoked earlier; that *indicates* that it is better not to smoke in the sense of doing something in order not get lung cancer. However, this association cannot directly be interpreted in *causal* terms as we cannot exclude third factors (hidden variables) operating in the background. One factor may be problems in handling stress that can lead people to smoke and at the same time make them more prone to lung cancer.

Another, simpler example illustrates the dilemma. If in the world-ski championship the Austrian skiers win one third of the medals, then we cannot infer from this fact that one third of the Austrian skiers are serious candidates for medals. They are not all the same, which is the basic assumption in the cancer example from before. They differ by third variables that might affect their success (disease above). This illustrates that cause and effect cannot simply be exchanged. Statistical methods are

specially designed to corroborate statistical evidence; they indicate in which cases it may be rewarding to search for contextual interrelationships but cannot provide a substitute for such research.

Knight's Distinction of Risk from Uncertainty

According to Knight “risk” and “uncertainty” differ by the character of probability: if the underlying probabilities are not known, then Knight speaks of “decisions under uncertainty”, if they are known then he calls them decisions of “risk”. This distinction clearly implies that risk excludes uncertainty, i.e., once the probabilities (and the impact) for the various events are estimated, uncertainty (or lack of any knowledge) is eliminated from the situation. Knight’s terminology has been so successful as – on the surface – it shifts the connotation of risk away from uncertainty (in the general sense) and makes the term “risk” seemingly more objective and thus more acceptable, which may be advantageous. The original quotation is:

The essential fact is that ‘risk’ means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; [...] It will appear that a *measurable* uncertainty, or ‘risk’ proper [...] is so far different from an *unmeasurable* one that it is not in effect an uncertainty at all. We shall accordingly restrict the term ‘uncertainty’ to cases of the non-quantitative type. (Knight, 1921, pp.19)

However, there are several objections. The use of “uncertainty” by Knight conflicts with its use in probability and philosophy. Philosophically, uncertainty (or chance) is opposed to determinism. Uncertainty is the general term designating (random) situations where a precise prediction of an outcome cannot be made with certainty, i.e., where probability has to be used in the prediction. Whether or not it is possible to get good estimates or relevant hypothetical values of the probabilities involved does not matter as the character of the probabilities (APT, FQT, or SJT) is not relevant. As the future is always uncertain, a risk is attached to actions relating to it.

A second objection is that Knight adopts a very naive usage of probability as – apart from textbook examples – probabilities are *not* known: even in games of chances there is always the dispute whether the conditions of a Laplace experiment with equal likelihood are in fact fulfilled in a specific case. Two other positions can also be stated: either (following von Mises) that probabilities are always only partially known and have to be estimated (FQT), and, probability does not exist in the objective world and is only part of our judgement on the world (SJT), as espoused by de Finetti. What we know is that the numbers are related to the models (e.g., an equiprobability model for a die yields 1/6 for each of the faces in the sense of APT probability) or estimates from past experience (FQT probability). Or, we know the values from elicitation from a person’s preference system (SJT probability).

A third objection is that from a subjectivist viewpoint the consistent choices of a rational person yield his or her subjective probabilities. Thus, while we have no precise values for probabilities (in FQT sense) we have probabilities (in SJT sense) and therefore the distinction between risk and uncertainty is not valid as noted by Friedman (1976, p.282).

Uncertainty has to be met by suitable procedures to get estimates or qualitative knowledge on the unknown probabilities. Of course, the status and the precision of probabilities that are used to calculate risks in the sense of (4) heavily influences the possible interpretation of this risk and its relevance for a discussion, especially if it relates to different stakeholders as is in the public discussion of safety of nuclear power or in public health issues. Of course, the more frequentist information about the underlying probabilities is available, the more objective seems to be the calculated risk. However, still there might be a dispute as different stakeholders can (and have to) exploit different information for their

purposes and they also “suffer” and benefit differently from the various actions to avoid the inherent risk (see Borovcnik, & Kapadia, 2011b).

Today, researchers defend the Knightian distinction between risk and uncertainty by saying that Knight did in fact link uncertainty to situations where markets collapse and so no subjective probabilities can be found (LeRoy, & Singell, 1987).

Knight was a key person in founding the Chicago School of Economics (with 3 Nobel Prize awards) and his ideas about risk and uncertainty have influenced many people. One example may illustrate how they express their ideas in Knight’s tradition:

“Risk: We don’t know what is going to happen next, but we do know what the distribution looks like. Uncertainty: We don’t know what is going to happen next, and we do not know what the possible distribution looks like. In other words [...] the future is always unknown — but that does not make it ‘uncertain’.” (Ritholtz, n.d.)

This view on the financial markets from a former trader, researcher and strategist, and now asset manager reflects a desire of eliminating uncertainty and arrive at objective figures for probabilities: there is risk but *no* uncertainty, and we have everything under control. This wording induces a misleading impression of the measurements taken or possible and reflects the fact that “experts” adopt a terminology that fits their purpose (rather than suiting their client’s interests).

Risk and Utility in Economic Theory

Risk is a key factor for economic activities and is the basic element for decisions on financial markets, which resembles the old paradigm of bets. In modelling economic activities, risk plays a central role therefore. If actions with the same expected utility are compared according to (4), the one with the larger variance is defined as more risky. Risky behaviour can then be described by properties of the utility function; to be risk averse means to prefer that one of two actions (with the same expected utility), which has less variance (in utility). The degree of risk aversion may be measured by the person’s willingness to pay to get one option over the other that is offered. The Arrow-Pratt measure of risk aversion is defined as $-u''(x)/u'(x)$ ($u(x)$ the utility function, which is presumed to be twice differentiable); the higher this value the more risk averse is the person.

Apart from theoretical frames like expected utility as the criterion for decisions, researchers also became interested in how people behave in comparison to normative theories and explain – mainly by reference to psychology – why they deviate from normative results. Kahneman and Tversky are famous for their experiments, in which they investigated how people really behave in simplified situations. People seem to replace utility by an asymmetric function and put much more weight on the tails of the probability scale, i.e., differences in probability have much more influence on actions when they relate to very small and very large probabilities as compared to probabilities in the middle (around 0.50). Tversky and Kahneman (1986) developed the so-called prospect theory to describe how actual behaviour deviates from theoretical models of rational behaviour under risk (for their experiments see below).

Analytic Investigation of Situations that Involve Risk

So far we have reviewed different concepts (1) to (5) to define risk and to discern risk from neighbouring concepts. The adverse events bear subjective components in two ways: their probabilities are not always judged by objective (FQT) information and the evaluation of impact may be biased by personal circumstances. A fuller comprehension of the mathematical concepts as well as psychological

factors may play a key role in processing the available information (in the extreme case even apparently objective information is just ignored or reinterpreted). We will deal with these issues separately below.

From a systems analytic approach, it does not suffice to define risk but it is essential to structure those situations that involve risk: which constituents build up a risky situation and how do they influence the way to deal with risk or the perception of risk? Borovcnik and Kapadia (2011b) refer to the following ingredients of a decision situation related to health: “the nature of the risk (and mathematical concepts like probability); the psychological matters involved; the type of situation (treatment, prevention) and information used; the people involved, their aims (purpose) and their inherent criteria”. These are general categories to describe risky situations.

Risk may involve only the decision maker who takes personal risks like travelling, smoking, nutrition, or personal health issues. A decision between alternatives is influenced by a subjective judgement of a balance between benefits and adverse effects. Decisions may also bring two or more stakeholders together, which typically is the case in (public) health. It is important to note that the stakeholders do not use the same information, or the same criteria to process the information, and they are also affected quite differently by the outcome. While a patient personally might suffer from consequences (no cure, side effects of treatment), the doctor is liable to the system to guarantee treatment according to the state of art. Other stakeholders in public health (including the media) still experience other impact and follow other optimizing criteria; for the dilemma resulting from the interests of the various stakeholders see Borovcnik and Kapadia (2011a, b). It is clear that the different stakeholders have to use different criteria to find or justify their decisions. Remarkable, however, is the inconsistent use of criteria by individual decision makers (see below).

The type of the situation is a further, often neglected influence factor on the perception of risk: there are “real” risks when an acute situation is there and some action has to be undertaken; think of an accident with a broken ankle. While there are several options, something has to be done to deal with the immediate situation. Sometimes there are virtual risks relating to the (far) future. The person might conceive cancer of the prostate. There are hazards lurking in the background but no direct link can be made to the adverse event. In public health, screening programmes have been installed for various diseases: mammography with breast cancer (see Gigerenzer, 2002, for general risk considerations in this case); or, PSA tests and biopsies with prostate cancer (see Sandblom, et al, 2011, for an attempt to evaluate the success of the programme). Another virtual risk is climate change. Whether or not there is a climate change, whether or not such a change will have adverse effects directly related to it, there is much effort and investment of money in an attempt to decrease the “identified” hazards for a climate change.

Statistical Notions of Risk

In this section we present statistical terms that – implicitly – deal with risk, or, vice versa, that build up a normative way to conceptualize risk. Risk is linked to the possibility and probability that a statistical test leads to a wrong “decision” and is identified as conditional probability. The Bayesian approach to conceptualize risk is linked to an expected loss argument. We also discuss the problems that arise from low probabilities as we have no data for such situations so that the decisions (and related risks) based on them lack empirical justifications.

Type I and Type II Errors

In the historical development of statistical tests, Neyman and Pearson used a decision theoretic framework. In the simplest case we have two probability distributions for a phenomenon and we do not know, which distribution really applies. We decide in favour of one after we see the data from a random sample from the “true” distribution.

Example from quality control. A production line fabricates items that have an average rate of defectives of $p = 0.04$ if the process is under control. If there is a disruption in the process, the proportion of defectives increases. We restrict the possible values from $p > 0.04$ to $p_1 = 0.10$; i.e., 10 % defectives are produced on average in this case. We have now two hypotheses; the null hypothesis H_0 : $p_0 = 0.04$ and the alternative hypothesis H_1 : $p_1 = 0.10$ and we search for a reasonable decision rule if we inspect a sample of $n = 100$ items from the current production. The situation is represented in Table 1.

Status	Decision	
	T_0 : do not reject (“accept”) H_0	T_1 : reject H_0
H_0 : $p_0 = 0.04$ Production is under control	Decision is correct	Type I error H_0 is erroneously rejected though it is true The production has been stopped yet everything was okay
H_1 : $p_1 = 0.10$ Production is out of control	Type II error H_0 is not rejected even though H_1 is true A serious disturbance of the production has not been detected	Decision is correct

Table 1. Risk of wrong decisions in a statistical test of the null hypothesis H_0 against an alternative H_1

There is a risk of wrong decisions in the sense of (1) as Neyman and Pearson formulated in their papers. The probabilities of wrong decisions (the adverse event) are conditional probabilities (conditional to the distributions in the null and alternative hypotheses). These probabilities of wrong decisions are often denoted by α and β and can be called risks in the sense of definition (3): $P(T_1 | H_0 : p_0 = 0.04) = \alpha$ and $P(T_0 | H_1 : p_1 = 0.10) = \beta$. The decisions T_0 and T_1 are based on the number X of defectives in the sample that is inspected. By the usual assumptions, X is binomially distributed with n ($=100$ here) and $p_0 = 0.04$ if H_0 and $p_1 = 0.10$ if H_1 is true. From here we can calculate α and β if we use a threshold of $X = 8$ for our decision, i.e., if we reject H_0 if $X \geq 8$ (T_1) and do not reject H_0 if $X \leq 7$ (T_0).

$P(X \geq 8 | p_0 = 0.04) = 0.0475$ and $P(X \leq 7 | p_1 = 0.10) = 0.2061$. Usually, the type I error is pre-set (often to 0.05. or 0.01) so that the threshold can be calculated and the type II error is dependent on the size of the type I error. In this setting we cannot obtain probabilities for wrong decisions, which are $P(H_0 | T_1)$ and $P(H_1 | T_0)$. In FQT view, hypotheses cannot have probabilities as there is no experiment for them (and therefore no relative frequencies); in SJT view, hypotheses have probabilities as any

(unknown) statement or event does have a probability but as long as the probabilities are not “elicited” the probabilities of wrong decisions are not known.

Errors of type I and II are interpreted by the following thought experiment (scenario; under laboratory conditions): If the null hypothesis always applies, then it will be rejected with probability α , that is, in $100\alpha\%$ of cases applying the test, the null hypothesis will erroneously be rejected. For our production line example this means that for $\alpha = 0.05$ the production will be stopped in 5% (on average) to search for specific causes but the production definitely is under control. If $\beta = 0.20$, then the production would not be stopped though the conditions of the alternative hypothesis apply. For the production line this means that – if the production is out of control and items are produced with an average rate of defectives of $p_1 = 10\%$ – the decision rule would not “recognize” that with probability 0.20 and let the production continue.

Such a primitive FQT interpretation is misleading as one never tests under laboratory conditions and it provokes an interpretation of α as an error rate that conveys something on the risk of wrong decisions. As we know that it is not an unconditional probability, we change the names and refer to the risk rather than the probability of a type I error. Taken literally, the pure FQT position leads to rationality gaps, which have been fiercely criticized in the controversy about the foundations of statistics (see Stegmüller, 1973, or Hacking, 1975, 1990). Thus, the type I error should be interpreted as qualitative figure that is needed to derive the threshold and the type II error is just a mathematical consequence of it. If it seems too large (as here), then more data should be sampled before the decision is made. The Bayesian approach would compare various decision rules according to expected loss in the sense of (4). We will see the approach at work in the example on journal copies later.

Fisher suggested to test a null hypothesis against data from a random sample and insisted on not to formulate a specific alternative hypothesis. Instead, he based his decision of rejecting the null hypothesis on the p -value, which is the probability of getting a sample statistic (say, a sample mean) as far as, or further (more extreme) from some reference value (which reflects the true population parameter). Fisher thought that the p -value could be interpreted as a discrepancy measure from the null hypothesis. Thus, the p -value would be interpreted as error probability of the test. However, as we have seen in our framework above, it reflects only the probability of erroneously rejecting a true null hypothesis, i.e., $P(T_1 | H_0)$ and not the probability of an error in decision making if we reject the null hypothesis, which is $P(H_0 | T_1)$. Thus it is misleading to associate the risk of a wrong decision with the p -value.

In the early attempts to formulate his ideas about significance tests, Fisher pre-sets the significance level α and rejects the null hypothesis if its p -value is lower than α . In comparison to the Neyman-Pearson test policy, this significance level coincides with the type I error. However, the significance test lacks information about specific alternatives and their related type II errors (β). The complement $1 - \beta$ is also known as the power of a test and conveys information about the probability that the test can “detect” a specific deviation from the null hypothesis if such a deviating hypothesis would actually be the true distribution for the investigated phenomenon. The power is a key concept to judge the risks associated to a statistical test.

While there are strong arguments for Neyman-Pearson’s view, in practice an alternative can often not be formulated like in the ANOVA case, when one has not yet a clear idea about the structure of the differences in mean between the tested populations and it is methodologically doubtful to extract hypotheses from the same data, which are used to test these hypotheses; or, in testing for independence,

where it is difficult to formulate specific hypotheses for dependence. Yet, Lehman (1993) tried to reconcile both approaches towards testing.

Risks with Low Probabilities

For events with very low probabilities, we have no data available and thus any FQT consideration is doubtful. What remains is an elicitation of subjective probabilities or the use probabilities in a scenario to investigate the issue on a “what if” basis. We illustrate the difficulty of getting reliable data for low probabilities by an example from Borovcnik (2012, p.22). Normally, an unknown probability is estimated by the related relative frequencies from a sample that is assumed to be random. This latter assumption is hard to meet, hardly questioned, and really difficult to check (a test for randomness lacks an alternative and therefore, a type II error cannot be calculated). Such an estimate has an inherent variability. Suppose, we want to investigate how we can get information for a probability p of 10^{-4} . If we base an estimate of p on a sample of “10,000 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”.

Another example is bovine spongiform encephalopathy (BSE, or mad cow disease), taken from Dubben and Beck-Bornholdt (2010, pp.64). There were times when all of Europe was in uproar because of news on new “cases” of BSE. Cattle with the disease are a hazard as those who eat their meat may contract Jacob-Creutzfeldt disease. Tests have been undertaken to evaluate the accuracy of diagnosing whether an animal has or does not have the disease. With the diagnosis test applied, a cattle gets the attribute positive T^+ (indicating the disease) or negative T^- (indicating that the animal is free of this disease). If the cattle has BSE, there is a “sensitivity” of (at least 0.99) that it will get a positive test result; if the cattle has no BSE, there is a “specificity” of (at least) 0.997 that it will get a negative test result. Both sensitivity and specificity are conditional probabilities: sensitivity = $P(T^+ | \text{BSE})$; specificity = $P(T^- | \text{no BSE})$.

These properties of the diagnosing procedure have been based on testing 300 infected cattle from British origin and from 1,000 animals from disease-free New Zealand. And in German laboratories all 1,300 biometric samples have been classified correctly, from where the properties above have been calculated by means of confidence intervals. However, put it the other way round and calculate the probability that 1,000 non-infected cattle are classified correctly as T^- given that the specificity is 0.997: to classify them all negative, there is – merely by chance – a probability of 5%. The extraordinary test result may thus be explained by mere randomness.

This reduces the 100% security of the laboratory test considerably. And, furthermore, remember that it is much more difficult to detect the still latent disease when no apparent traces of it are there than it is to confirm the disease (by the same diagnosing test) of a full-blown BSE cattle. That is, under conditions of mass application the careful procedure of the laboratory test may not be upheld. A cattle classified positive need not be infected; we cannot claim that with certainty. Likewise, a cattle classified negative need not be disease-free. To calculate the probability that a cattle with diagnosis T^+ in fact has BSE, we need also the prevalence of BSE in the population of German cattle. This is an unknown number. While there are subjective estimates, there is no objective number.

There were 331 cases of positive tests in Germany from January 2001 till June 2004 when 9,747,738 cattle have been tested. That means the prevalence is less than 3 out of 100,000. It could well be that this number is much lower and all the positive test results have been false positives. We will

never know. This illustrates drastically the problem with hazards that have a low probability to cause harm. As the impact is high (human beings may contract a deadly disease), societal rationality is put under enormous stress.

Another issue arising from low probabilities is that we underestimate cumulative risk. Suppose there is a personal risk in climbing. For one trip a risk might be calculated in the sense of (3) of 0.0001, which is negligible. For 2,000 trips the probability of not being involved in the adverse event attached to it is calculated to $0.9999^{2,000}$ whence the cumulative risk is 18%. What if the initial risk is 0.0003 (which is still a negligible number)? The cumulative risk rises to 45%. Who would have expected that effect? The difficulty in understanding such calculations means that the cumulative risk is neglected though it is essential.

Psychological Aspects of Risk

Risk involves the probabilities of several outcomes and their impact. The probabilities may attain a more subjective (SJT) or a stronger objective (FQT) character. The impact can be measured by “money” or utility; the latter has a strong subjective component. It is interesting to know whether people share similar forms of utility and how far they are consistent between different choices they make. A further key issue is whether people behave rationally or are influenced by psychological factors. The criteria people choose might be dependent on their perception of risk, which may be influenced by context, by low probabilities, or by high impact. Surprisingly, there are arguments to change the criterion for decisions depending on whether a decision is made once or made repeatedly. We start with some famous experiments of Kahneman and Tversky.

Probability and Utility

Kahneman and Tversky conducted many experiments with adults on understanding probability and risk and showed how inconsistent people are with regards to uncertainty. They characterised their behaviour as misconceptions and formulated a terminology: representativeness, availability, anchoring, etc. The examples used to explore people’s behaviour resemble the following one though some are more complicated (Kahneman, & Tversky, 1979; Tversky, & Kahneman, 1981).

Experiment 1: Win [in \$]	Option a ₁	Option a ₂
Potential future	1,000	2,500 with $\frac{1}{2}$ 0 with $\frac{1}{2}$
Experiment 2: Win [in \$]	Option a ₃	Option a ₄
Potential future	–1,000	–2,500 with $\frac{1}{2}$ 0 with $\frac{1}{2}$

Table 2. Type of experiments, Kahneman and Tversky (1979) used for their studies

In experiment 1 (see Table 2) people prefer \$ 1,000 for sure even if the second option has a value of 1,250 while in experiment 2 people prefer the second option and seek the risk rather than the sure loss of \$ –1,000. In winning situations people are risk averse. People prefer \$ 1,000 for sure even if the second option has a value of 1,250. On the contrary, in losing situations people are risk seeking: they prefer the second option with a value of –1,250 and seek the risk rather than the sure loss of –1,000.

“Problems [...] demonstrate that outcomes which are obtained with certainty are overweighed relative to uncertain outcomes. In the positive domain, the certainty effect contributes to a risk averse preference for a sure gain over a larger gain that is merely probable. In the negative domain, the same effect leads to a risk seeking preference for a loss that is merely probable over a smaller loss that is certain. The same psychological principle – the overweighting of certainty – favours risk aversion in the domain of gains and risk seeking in the domain of losses.” (Kahneman, & Tversky, 1979, pp. 268)

In developing prospect theory Kahneman and Tversky (1979) found that the decisions observed in their experiments are incompatible with expected utility theory that has been proposed as normative theory of behaviour in view of risk by Neumann and Morgenstern (1953) and widely used as a model to describe economic behaviour (Arrow, 1971). In 2002, Kahneman was awarded a Nobel Prize for his work on the psychology of decision making (of which prospect theory is a foundational part).

With the usual utility functions (concave for positive and convex for negative values of money), such behaviour is inconsistent. That implies also that people have to act in the experiment as if it were real which is doubtful as in virtual games decisions are different from the hard consequences of real life where no switch-off button exists. The method also requires that people can disentangle the perception of probabilities and utility considerations; however, the perception of any probability is always linked to its impact (the probability of an event with a severe negative impact is often highly over-estimated). Furthermore, Kahneman and Tversky only described the phenomenon that has since become famous: people are risk-averse in winning situations and are risk-seeking in losing situations. They did not explain this behaviour by their analyses.

We will twist the representation of the experiments by a reformulation, which changes the character of the winning situation to a losing situation and vice versa. By this we show that the representation of a situation is not unique and implications of the behaviour cannot easily be drawn. Furthermore, we will explain the behaviour, which seems obvious in the new format. And, the result can be reconciled with utility with a somewhat different utility function, which has not yet been considered but seems relevant (one that is bounded from below and the lower bound is reached very “soon”) especially if one investigates ongoing political and financial decisions around the rescue of the Euro since 2008.

Experiment 1 can be formulated differently. People think of \$ 1,000 for sure as they have it already with option a_1 . The decision for a_2 then appears as relative to the status quo as potential losing situation as they imagine losing \$ 1,000. Instead of comparing a_1 and a_2 , they compare a_1 and a_2^* . This sheds doubt on Kahneman and Tversky’s hypothesis “in winning situations people are risk averse” (Table 3) and “in losing situations people are risk seeking” (Table 4).

Experiment 1 [*]	Option a_1	Option a_2^*	Option a_2
Potential future	1,000	1,500 with $\frac{1}{2}$ –1,000 with $\frac{1}{2}$	2,500 with $\frac{1}{2}$ 0 with $\frac{1}{2}$

Table 3. Reformulation of experiment 1 from a winning to a losing situation

The possibility of winning an extra amount of only \$ 1,500 is now related to the additional risk to lose the 1,000 and go down to zero! It may not pay to risk the 1,000 for the small extra win of 1,500. The situation may thus be perceived as losing situation. A person that has already 1,000 has to be paid

(much) more to take a risk to lose 1,000. The more a person already owns, the more this person has to be paid to seek the risk.

Experiment 2*	Option a_3	Option a_4^*	Option a_4
Potential future	-1,000	-1,500 with $\frac{1}{2}$ 1,000 with $\frac{1}{2}$	-2,500 with $\frac{1}{2}$ 0 with $\frac{1}{2}$

Table 4. Reformulation of experiment 2 from a losing to a winning situation

In the same way we can reformulate the situation of the second experiment and regard the option relative to the status of having already lost 1,000. The situation is now a winning situation. People who already have debts will take more risks to balance them. In our own experiments with students for these tables, we have found that the hypothesis of Kahneman and Tversky is not supported; furthermore, we have also found that people are markedly influenced by their own disposable income.

Inconsistencies and Different Perceptions

Behaviour may be influenced if utility of money is involved, if payments are much bigger, or if initial situations are perceived (presented) differently. It might be the same people who avoid a_2^* and seek a_4^* , i.e., avoid winning an extra 250 (as expected value) but seek to lose another 250. It is not inconsistent as it compares the behaviour of the same person in two different settings (one with already a reasonable amount of money, the other with an unbearable loss). That changes the alternatives at stake and influences their behaviour switching their decision. It is important to note that judgements can only be seen with the related alternatives at stake. Thus, it is not feasible to compare a_2^* and a_4^* directly.

One can now change the alternative involved to trace the consequences thereof. If we consider the offer to win a very high (but improbable) prize compared to a small amount to pay (as in the state lottery), we will see that a lot of people are risk-seeking in this winning (in the sense of Kahneman, & Tversky) situation. On the other hand, if a potential (but improbable) high loss can be “balanced” by a relatively small amount of money as in an insurance contract (which is very popular nowadays as people try to get a contract on nearly everything to eliminate uncertainty) in losing situations, we see evidence that they are risk-averse in losing situations (again in the Kahneman, & Tversky sense).

The two latter findings about the ubiquity of games of chance (and betting) and the popularity of insurance contracts show a distinct deviation from what Kahneman and Tversky claim and gives our reformulation more justification. Winning or losing is not a predicate for a situation that can be unambiguously granted. It is always a matter of reconstruction of the situation involved. Rather than to regard a situation as winning or losing situation, one might focus better on the original status (initial wealth) of the test person. If a person earns 1,000 per month or 5,000 it does affect the decision; it might influence it much more than the various amounts that are at stake; similarly, if a person has great debts already with no perspective to fulfil them, this person might be willing to accept risks that other persons would find completely irrational; indeed that is the basis for some gamblers. Also, these “facts” indicate that people do not focus so much on probabilities but much more on the impact related to the potential outcomes.

The preceding considerations also indicate that it is difficult to separate between the perception of impact and probability of future outcomes. Personal criteria come to the fore. In a winning situation, beliefs like “I deserve it (to win the prize and lead a happy life)” let people gamble (seek the risk) and wait for “God’s decision”. In a losing situation, they are willing to pay an amount of protection money

to ease out “God’s enagement” to avoid not only the “accident” but also the negative consequence related to it that goes beyond the financial loss. The personality is a further factor that influences decisions under uncertainty; it is possible that women – on the average – have a distinct risk profile from men. All these considerations reveal how difficult it is to weight the evidence rationally. Gigerenzer (2007) investigates the effect of gut decisions on the quality of decisions: first, gut decisions are popular; second, they are not necessarily worse than careful analysis of all givens; third, it might be good to support quicker decisions by checklists about what to consider. A mixture between analysis and gut decisions might be superior to pure analytic decisions not only because of the undue delay in decision by careful analysis including collecting all the necessary information.

A final example about personal criteria refers to the three-door problem with the two goats and the car behind the doors where the candidate can freely choose one. After the candidate has made the choice, the moderator opens one of the remaining doors to show a goat and offers the option to change the first choice, i.e., to switch to the still closed door or remain with that door selected first. Borovcnik (2012) discusses various feasible strategies to analyse the problem and whether it is advantageous to change one’s first choice. Of course, by switching the candidate increases the chances to win the car from $1/3$ to $2/3$. However, with all the explanations, people – in the majority – stubbornly remain with their first choice even if they are fully capable to understand the solution. The discussion is signified by fierce emotional expressions so that one may ask for the reason behind. One simple explanation to refute all mathematical solutions might be that the candidates simply compare the situations afterwards and relate a strong impact to these: to stay with the first choice and to lose will leave them as unlucky; they simply had not a “fair chance” to win the car; to lose the car if they had chosen the right door and to switch to the remaining door will leave them with the irony of friends and with the feeling that they should not challenge God’s will. They do not compare risks in the sense of (3) or weighted risk in the sense of (4); no, they simply compare the impact of their decision. And the impact of losing by their action and being responsible for this is by far the worst they can imagine. Therefore, they defend their first choice against all odds.

Idiosyncratic criteria, gut feelings, fear of high loss in the worst case, eager, responsibility for what one does (as compared to bad luck) always overlays a rational approach; it starts with hindering to separate the impact from the (small) probability of occurrence; it biases the perception of the (small) probabilities of adverse events; and it increases the factual impact of the adverse event and focuses on the worst case letting any weighting seem irrational and unfeasible.

The Logic of Repeated Decisions and One-Off Decisions

In a statistical variant of experiments 1 and 2, a candidate has to decide repeatedly (1,000 times) between the presented options. Instead of performing the experiment, we just tell what we would do and give a rationale for it. What would the reader decide in the situations presented in Table 5? Again, utility may play a role but as the payments are so small, more or less the amount counts (this argument might not apply if the single amounts were higher). We decide for the risky option b_2 in the winning situation of experiment 1^s and for option b_3 in the losing situation of experiment 2^s. Thus we seek the risk in the winning situation and avoid the risk in the losing situation (terms applied in the sense of Kahneman, & Tversky), which is different from our behaviour in the non-statistical variant of the experiment. We will not transform the situation relative to the “left” side but give a statistical argument that should suffice.

Experiment 1 ^s : Win [in \$]	Option b ₁	Option b ₂
Potential future	1	2,5 with $\frac{1}{2}$ 0 with $\frac{1}{2}$
Experiment 2 ^s : Win [in \$]	Option b ₃	Option b ₄
Potential future	-1	-2,5 with $\frac{1}{2}$ 0 with $\frac{1}{2}$

Table 5. Statistical variant of experiments 1 and 2: 1,000 games have to be played; each requires a decision between the two given options

To win an amount of more than 1,000 in the first experiment, we need to win in more than 400 single situations. The probability can be easily calculated from the binomial distribution (with $n = 1,000$ and $p = \frac{1}{2}$) to a figure, which is close to 1 ($1 - 10^{-10}$). We have a probability of at least 0.51 to win 1,247! For the losing situation now the results are just reversed: we have a probability of nearly 1 to lose more than 1,000 and we lose 1,247 with a probability of at least 0.51.

It makes a difference in terms of the success of the strategy used if one has a one-off decision or decides similar cases repeatedly. What is good in a one-off decision may be bad for the repeated decision. In other words, to speak with an illustrative context: if one sells a house one time, the relevant criteria differ from those if one sells houses every week like a professional agent; this is not because of the impersonal impact but because of doing it repeatedly; of course, a professional agent has additional criteria for decisions, too.

Three Paradigmatic Examples of Risk

In the following three examples, we present the genuine character of situations under uncertainty and discuss various useful strategies. The criteria used are not always defensible and they have their relative merits and disadvantages. Of course, the decision based on a specific criterion will depend on it and possibly switch if the criterion is changed. If probabilities are involved in the decisions, their values can change even in the same problem for the various stakeholders involved. There is no unique view of a problem; rationality depends also on the position of a stakeholder in a decision. Decisions under uncertainty – if not “played against nature” may also be interpreted as an exchange of certainty and money between the stakeholders as is typically done in an insurance contract, which is a generalization from bets within games of chance.

Example 1. Insurance – Exchange Money for Certainty

We will embed matters in a decision with two stakeholders that exploit different types of information for their decision. In any insurance contract, two partners mutually exchange money and the status of uncertainty. For a full-coverage insurance for the car (for one year), the insurance company gives up its position of certainty (no loss) and offers to pay the potential costs from an accident while the client pays a certain amount (the premium) in advance in order to leave the situation of uncertainty (of an accident) and reach a status of certainty over the financial cost of an accident. A similar situation occurs in any bet in the state lottery, bets in sports, or with option contracts in the financial market.

If both parties apply the same principles then the question arises how can the contract be advantageous for both – it seems like a paradoxical situation. However, both stakeholders have their own viewpoint and use different criteria that fit their circumstances better. The insurance company has many contracts and thus can use the frequencies of accidents from past statistics to estimate the underlying probabilities. Since the company accumulates assets, it is free of utility considerations. The client, however, is a unique person with different habits (driving skills, driving regions, exposition to special risks, etc.), which emerge into a personal judgement of probabilities of accidents. Furthermore, the client does not have large financial assets so that utility considerations of money become relevant, e.g., can one afford the cost of another new car within one year in the worst case of an accident? It is interesting that states normally have no insurance for their cars. They have a larger fleet of cars and they have a different financial background. We discuss a crude model for the potential future of one year considering only a total wreckage or no accident below (Borovcnik 2006; see Table 6).

Cost (€)		a_1 = Insurance yes	a_2 = No insurance
Potential future	t_1 = No accident	1,000	0
	t_2 = Total wreckage	1,000	30,000

Table 6. A crude model for the insurance contract

The insurance company may base its model on money and an estimate of the probability for the damages (and related payments) by past frequencies of events. For data of 2% for the wreckage the calculation is: $30,000 \cdot 0.02 + 0 \cdot 0.98 = 600$ plus ... expenses and profit. The car owner has to find his or her personal probabilities and take utility of money into consideration. Without utility, a so-called break-even point can avoid the difficult process of eliciting the personal probabilities. The break-even here is for odds of 1 : 29 for the total wreckage as in this case the decisions for the insurance or against it each have the same cost, namely 1,000 so that the person (not considering utility) would be indifferent between the two actions. If the model considers also smaller accidents (parking damages, e.g.) taking out the insurance becomes more attractive.

Example 2. Using Probabilities for Optimization of a Decision under Uncertainty

We illustrate how a person willing to model uncertain future outcomes by probabilities may optimize a current decision. Suppose the demand D for a journal is uncertain. We model it – for reasons of simplicity – by the following discrete probabilities p_i . Cost of production $C(a_j)$ (in €) is dependent on units a_j which are printed, as given in Table 7. The selling price of one issue of the journal is € 1.60. How many units should be printed if you have a choice to print 1, 2, ..., or 5 thousand copies?

Demand d_i	1,000	2,000	3,000	4,000	5,000
Probabilities p_i	0,40	0,30	0,20	0,06	0,04
Copies a_j	1,000	2,000	3,000	4,000	5,000
Cost $C(a_j)$	2,000	2,200	2,400	2,600	2,800

Table 7. Probabilities for demand for the journal and cost of copies

Calculation of the Expected Profit of a Single Decision. The cost is determined by the decision but the income remains subject to randomness. We might calculate its expected value in the sense of (4) in order to judge the present decision. We get a monetary value but cannot interpret it in isolation; we will have to calculate the expected profit also for the other possible actions and then compare the values.

Decision a_2 : 2,000 copies with cost $C(a_j) = 2,200$			
Demand d_i	Probabilities p_i	Gross income	Net profit
1,000	0.40	1,600	-600
2,000	0.30	3,200	1,000
3,000	0.20	3,200	1,000
4,000	0.06	3,200	1,000
5,000	0.04	3,200	1,000
Expected profit (a_j)			360
Maximum loss (a_j)			-600

Table 8. Calculation of the consequences of a specific decision

The first entry of the *net profit* in Table 8 is -600 as when only 1,000 copies are sold, then the net profit is $1,000 \cdot 1.6 - 2,200 = -600$. It is important to note that an action cannot be judged in isolation: how could we interpret the expected profit of 360, or a maximum of loss of -600? A rational judgement requires the comparison of alternatives. Thus, we repeat the analysis for alternative numbers of copies (1,000 to 5,000) and arrange the results of computations in a combined matrix of net profits below (Table 9).

Comparing Different Decisions. We compare our options with respect to the number of copies according to the criterion of expected profit (rather than money we could also use utility of money).

Demand d_i	Net profit of decision: Number of copies a_j					Probabilities p_i
	1,000	2,000	3,000	4,000	5,000	
1,000	-400	-600	-800	-1,000	-1,200	0.40
2,000	-400	1,000	800	600	400	0.30
3,000	-400	1,000	2,400	2,200	2,000	0.20
4,000	-400	1,000	2,400	3,800	3,600	0.06
5,000	-400	1,000	2,400	3,800	5,200	0.04
Cost $C(a_j)$	2,000	2,200	2,400	2,600	2,800	Price per issue 1.60
Expected profit (a_j)	-400	360	640	600	464	
Maximum loss (a_j)	-400	-600	-800	-1,000	-1,200	

Table 9. Matrix of net profit depending on the decisions and the actual demand

The second column of Table 9 contains the random variable *profit* for the option of 2,000 copies, which was already displayed in Table 8. Other entries are calculated analogously. The probabilities for the entries in the second row are derived from those for the demand (they are contained in the outmost right column). According to the options we compare five different distributions for the profit; we could write them in separate tables and draw their bar graphs. As the probabilities are the same, we write the distributions compactly in the profit matrix.

It is obvious that no one would be willing to decide for option 1,000 as – whatever the demand will be – it will lead to a loss of -400. This reminds us to a principle of avoiding a sure loss (which is a basic principle in SJT). The option 2,000 delivers a positive expected profit of 360; however, it can lead

also to an even higher loss of -600 as compared with the decision for 1,000 copies. This reflects a basic property of improving decisions under risk. Rarely can one find decisions, which are better throughout (whatever will happen), but it is easy to rule out inferior decisions. However, the remaining (admissible) actions cannot be compared to each other without a further criterion and what is the better decision depends on the criterion used. To improve a situation in one respect (to have a higher expected net profit) is accompanied by the risk of higher potential losses. One may even speak of an invariant in human life as seen from a general philosophical perspective on risk. The option 3,000, which yields an expected profit of 640, is better. However, it bears the risk of a loss of -800 (if demand is only 1,000, which has a probability of 0.40). It turns out that option 3,000 yields the maximum expected profit (640) and is – in the present model – the best decision.

The decision depends on the criterion used. If the criterion for the decision would be to minimize the maximum possible loss, then option 1,000 copies would be the best but this is not a feasible option at all. A minimax principle (minimizing the maximal possible loss) thus may lead to nonsensical decisions. Again, a general comparison is that fearing the maximum loss will often end up with a poor decision. If probabilities are modelled for the demand for copies of the journal (e.g., from past relative frequencies), then an expected profit criterion may be applied. Generally, this leads to better solutions but bears the risk of somewhat higher losses. In this way the Bayesian view would use risk in the sense of (4). The distribution of the demand is usually the result of validating a prior distribution on the demand by empirical data via the Bayesian formula and has a SJT character. Instead of basing the calculations on money, we could also attribute a utility to the different amounts of money.

Example 3. Medical Diagnosis

Medical diagnosis is a natural context to introduce conditional probabilities and reflect on the possible errors of making decisions. Suppose we have data from a group of 1,000 persons (see Table 10) about the cross-classification of their joint status of a disease D and the result of a biometric variable that is used to classify them as positive (T^+) or negative (T^-). The question is whether the result of the diagnosing variable can be used for diagnosing the disease or not.

Diagnosis	Status of disease		All
	D	Not- D	
T^+	9	99	108
T^-	1	891	892
All	10	990	1,000

Table 10. Data from an investigation in the disease D and a diagnosing method to detect D .

We imagine that each of the persons corresponds now to a ball that has two markers, one for the status of the disease and the other for the result of the diagnosis. We put all the balls into an urn and draw from it randomly. We get the following probabilities, which are easily read off Table 10:

$$\text{a. } P(D) = \frac{10}{1000} = 0.0100; P(T^+) = \frac{108}{1000} = 0.1080, P(D \cap T^+) = \frac{9}{1000} = 0.0090.$$

$$\text{b. } P(T^+ | D) = \frac{9}{10} = 0.9000 \text{ as compared to } P(T^+ | \text{Not} - D) = \frac{99}{990} = 0.1000.$$

$$\text{c. } P(D | T^+) = \frac{9}{108} = 0.0833 \text{ as compared to } P(D | T^-) = \frac{1}{892} = 0.0011.$$

We will use the name *prevalence* (incidence) for the probability of the disease. We can speak of the probability of a positive diagnosis to be 0.1080 in a. In b., we compare the probability of a positive diagnosis within “subgroup” D (0.9000; first column) to the analogue probability in “subgroup” Not- D (0.1000; second column) and state that the likelihood to get a positive result is 9 times larger in the group who have the disease D than in the group of people who do not have the disease. We can also give a name to the probability of a positive result T^+ in group D : this is the *sensitivity* of the diagnosing method. Its complementary probability is linked to an error of the diagnosis; here we have an error rate of 0.1000; i.e., on average we will not detect 10% of those who have the disease (as they get a negative result T^-). On the other hand, we can state that the diagnosis T^+ is wrongly attributed to a person without the disease with probability 0.1000. The complementary probability is called the *specificity* of the diagnosis and it reflects the “reliability” that a person without D gets a correct negative result.

The context illustrates that diagnosing for the disease (or for the absence of the disease) is linked to risk. We can make two errors in deciding that a positive person has the disease and a negative person does not have the disease D . We can wrongly classify a person as to have the disease (first row, second entry) and we can wrongly classify a person not to have the disease (second row, first entry). The size of the errors, i.e., the probability of the errors characterizes the quality of the diagnosing procedure and whether the underlying biometric variable should be used for diagnosing D . Usually, the prevalence is estimated and the sensitivity and specificity are checked by a laboratory study of cases which are classified by other methods so that we certainly know of these persons their status of D . If we reformulate the diagnosing problem as a statistical test of the null hypothesis “to have not D ” and the alternative hypothesis “to have D ”, then the false positives (1 *minus* specificity) becomes the error of type I while (1 *minus* sensitivity) becomes the error of type II.

Elementary Approaches to Probability and Risk

Probabilities, especially conditional probabilities are difficult to perceive and hard to estimate and interpret in a specific context. Also, such probabilities reflect that we explore a real situation only via models that need not fit to it. There are interesting proposals to simplify the probabilistic data and to visualize them so that the information is more readable.

Gigerenzer’s Tables with Natural Numbers

Gigerenzer (2002) has suggested the use of absolute numbers instead of (conditional) probabilities. All analyses are then based on these “natural frequencies” and are simplified to divide two numbers (in the sense of APT; see also Table 10). His research team has also investigated, which visualization is better to obtain and understand the solution, which is usually derived from calculations that are rarely understood (and are based on the Bayesian formula).

Suppose the data for the screening scheme for the detection of breast cancer are the following (Table 11; Gigerenzer, 2002): Of women between 40 and 50, 0.8% conceive cancer of the mammae yearly; we call this prevalence (incidence) of the disease. From those who have breast cancer, 87.5% are detected by anomalies in the mammogram (this is the sensitivity of the diagnosing procedure), i.e., diagnosed positive; from those who do not have breast cancer, 7% are also diagnosed positive indicating falsely that they have cancer (the specificity is 93%). In a probabilistic framework, a woman is randomly selected from the population and undergoes the mammogram. After she has a positive result, the question is what is the conditional probability to have cancer of the breast. An analogue question can be put after the mammogram is negative: what is this person’s probability to have no cancer of the breast

given that the mammogram is negative? The answer can be obtained by the Bayesian formula, which is quite formal. Gigerenzer suggests transforming all (conditional) probabilities to natural frequencies (expected values) for a group of 1,000 or more persons.

	Status		All
	D	Not-D	
T^+			
T^-			
All	8		1,000

	Status		All
	D	Not-D	
T^+	7	70	77
T^-	1	922	923
All	8	992	1,000

Prevalence (Prior): $P(D) = 0.008 = 0.8 \%$; Sensitivity: $P(T^+ | D) = 0.875 = 87.5 \%$;
 Specificity: $P(T^- | not - D) = 0.93 = 93 \%$.

Table 11. Transforming (conditional) probabilities to natural frequencies by expected numbers in a statistical village

We round off the value of 922.56 to 922 for simplicity. The other data can just be filled by the usual side constraints. From the table of natural frequencies we can calculate any probability related to the diagnosis. If we restrict the selection to the first row, we read $P(D | T^+) = \frac{7}{77} = 9.09 \%$, if we restrict the selection to the second row, we obtain $P(not - D | T^-) = \frac{922}{923} = 99.89 \%$. Remarkably, the positive diagnosis has an unexpectedly low conditional probability for the disease, because of its low prevalence. The advantage of Gigerenzer's approach is that the formalism of Bayes' formula is circumvented and the final result is more convincing as only 7 from 77 positive diagnoses relate to women with breast cancer. In fact, Gigerenzer suggests that the expected numbers are arranged in a tree diagram instead of two-way tables as was done here.

Spiegelhalter's Icon Arrays

Spiegelhalter (2014) uses natural frequencies and presents each of the persons (units) in the statistical village by an icon. In his icon array he systematically compares the expected numbers of two scenarios, one with the screening scheme applied and the other without over a longer period (see Figure 1). This allows judging the benefit of the screening as compared to a decision for not undergoing the screening regime. Data in Table 12 display the expected numbers and illustrate the effect of PSA screening and digital rectal examination; numbers are for men aged 50 years or older, not participating vs. participating in screening for 10 years (1,000 men in each group).

Category of outcome	1,000 men without screening	1,000 men with screening	Icon used
Men dying from prostate cancer	8	8	●
Men dying from other causes	192	192	⊗
Men diagnosed for prostate cancer and treated unnecessarily	–	20	●
Men without cancer that got a false alarm and a biopsy	–	180	⊗
Men that are unharmed and alive	800	600	○
All	1,000	1,000	

Table 12. Expected numbers of diverse outcomes in a ten year screening scheme applied to 1,000 men as compared to the outcome of no screening (exact figures are slightly smoothed here)

1,000 men without screening:

1,000 men with screening:

Figure 1. Icon array to illustrate the expected outcome for the prostate screening system as compared to no screening (from Spiegelhalter, 2014)

In the usual format one has to learn to read the information (and the formalism). Many formats in scientific communication hide more than they clarify the inherent information. Thus, there is an urgent need for elementary approaches. However, there are also disadvantages: The basic assumption in the approach of natural numbers is that people are all alike. On the one side there is a psychological barrier to accept such an averaging view. On the other side, there are not 1,000 men who are like me (as is suggested in the icon array). The data and the derived results when enhanced by the visual representation appear as factual and true. The key point is that whole numbers or visual representations of them can appear to give more validity than is appropriate as models and imprecise information are used. Both the probability of prostate cancer (the prevalence) and the sensitivity and specificity of the diagnosing procedures used in the screening regime are crude estimates for the underlying probabilities.

Odds and Probabilities

We can use odds for calibrating (tuning) our personal probability in the sense of SJT. Odds are the proportion of the probabilities that “ E occurs to E fails to occur”, i.e., $P(E):P(\bar{E})$. Odds for a six on a die are 1 : 5 and for a head on a coin, the odds are 1 : 1 (fifty-fifty). From odds, the probability may be recalculated by $P(E) = \frac{p}{p+q}$ so that $P(six) = \frac{1}{1+5} = \frac{1}{6}$.

In the diagnosing example from above we have prior odds for the disease of 8 : 992 or roughly 1 : 120. The probability of a positive diagnosis T^+ is 7/8 in the subgroup with D , and is only 70/992 in the subgroup without the disease D whence we state that the data likelihood for a positive diagnosis T^+ equals 7/8 : 70/992, or roughly 12 to 1; this expresses the fact that a positive result has a probability that is 12 times higher in the subgroup of women with breast cancer (D) than in the breast cancer-free group ($\bar{D}$). The formalism of the Bayes’ formula can be reduced to multiply prior odds and data likelihood (see Borovcnik, 2012).

We represent the odds by fractions and multiply as if the factors are scaling factors for the odds (in the same way as scaling factors are multiplied in copying a sheet of paper repeatedly, scaling the intermediate copy up or down, to obtain the final scaling factor): $\frac{1}{120} \times \frac{12}{1} = \frac{1}{10}$, which yields posterior odds for D given the positive result T^+ . The simplified posterior odds of 1 : 10 yield a posterior probability for D of 1/11, or 0.0909. The advantage of this calculus with probabilities is that we can identify the low prior odds of the prevalence to cause the surprisingly low probability to have breast cancer after a positive result. The diagnosing procedure per se is not very good but the data likelihood of T^+ increases prior odds by a factor of 12, which is not bad at all.

Evaluating and Sharing Risks between Different Stakeholders

Apart from personal risks, one may differentiate risks to the field where they arise (gambling, investments, etc.) or to those who are involved: single persons, or the society as a whole. In the latter case we might speak of societal risk. Climate change is a classic example of societal risk. Risks in medicine are harder to classify as there are single patients who are at risk but there is also an institutional side involved. The institutional side has its own goals which may differ from that of each individual.

Risks Shared between Stakeholders of Different Levels

For example, in public health there may be a discussion whether children should be vaccinated against measles or not. The health system is concerned about public order and is responsible to prevent an epidemic outbreak of the disease. It can build a frequentist scenario and weigh the different possibilities by probabilities and calculate the risks involved in wide-spread vaccination (possibly with a duty to partake) and reduced vaccination (no public financial support, no advertisement and counselling of parents). The individual parents have their subjective probabilities (own experience with the disease, strength of immune system, alternative ways to support health and cure in case of a disease, etc.). They cannot rely on the frequentist probability of the public health system, and the impact is hard to judge: the possible impact of the disease and the possibility of a bad consequence of the vaccination itself. While the parents stay with the personal consequences in either case, the doctor can be made liable for not informing about the possibility of the vaccination (or its negative side-effects). The pharma industries are a further stakeholder in this case, which has completely different financial interests and goals. We

can see from this description that such decisions cannot be prescribed but have to be made personally, with all the assistance of public information that relates to the current state of medicine. For the different situations of the stakeholders in case of the recent HPV vaccination (that is intended to protect from cervical cancer) see Borovenik and Kapadia (2011a).

Medical doctors enhance and increase their role by warning on health risks. With these warnings they are always on the safe side. If the patient follows their advice and the negative consequences do not occur, their concern seems justified; but if the patient still suffers from some consequences, there may be many other explanations and still the doctor's concern proves right. If the patient does not follow their advice and gets "ill", this seems again to be a confirmation of the doctor's concern; if the patient does not experience the negative outcome in this case, there will yet remain a reminder (you were only lucky!).

Societal Risks – Standards of Science and Society

In technological settings, the evaluation of risks has become a driving force to develop related concepts. While it seems desirable to keep technological risks as small as possible, economic reasons would be a "natural" limitation. And, there are no absolutely safe technologies so that the key question becomes whether we can reasonably decrease the risk.

When risk considerations are applied to complex systems (climate, ecosystems, world economy, etc.), the systems contain so many components so that their interactions become practically unpredictable. Whether climate change really establishes a risk remains doubtful for many people. The probabilities for the various developments are hard to assess and may be changed by time, technology, and economics. Many attempts are undertaken to simulate models (scenarios) on the basis of assumptions (with the well-known effects of flooding, increase of hurricanes, etc.) lacking serious attempts to estimate the probabilities of certain quantitative effects. The models resemble scenarios (what happens if ...?) and cannot claim the status of a precise description of the real world as too many weakly supported hypotheses play a key role in their considerations.

This example reveals a key problem in presenting scientific results to the public. Standards that are viable within research may not be applicable for that generalization especially as the stakeholders have their own probabilities and are affected by the consequences completely differently. And there are dynamic developments involved that cannot be influenced further. For example, once there is a public decision for the energy use of atomic power, there are industries involved in the business and make their profit while on the other hand the public has to balance if an accident happens (as in Fukushima 2011, with the potential harm on health but also financially to cope with the damage).

The terms "objective" risk (when objective probabilities are involved) and "subjective" risk (when subjective probabilities are involved) are misleading, especially if the risk has to be shared between stakeholders of different levels. The calculated risks are only a support for finding a common decision but never can be used in the way that decisions are executed. Apart from the risk as it is per se, there is a perception of risk that is not based on (well-calibrated) subjective probabilities but more or less on intuitive short-cuts of such a risk calculation, which is based on beliefs and gut feelings.

When scientific results should be applied to other areas, especially to public decisions in society, we have to clarify the standards of evidence. In sciences, one essential strategy is to find hypotheses of specific relations between variables by rejecting a null hypothesis (that there are no such specific relations) and the type I error (of falsely rejecting the null) becomes vital. However, in applications a specific hypothesis has to be applied and the type II error of wrongly not rejecting the null hypothesis

(of no special relations) becomes the nub of considerations. What if this type II error is very small? Then it might be very easy for an alternative hypothesis of special relations to get acknowledged by rejecting the null. We have dealt with these types of errors in a separate section earlier as they form part of the inner-mathematical approach towards risk within hypothesis testing.

Here an example illustrates matters. If a null hypothesis is rejected (because of a moderate or larger type I error), then one would establish a special relation between variables and use such a model for further action. What if this superimposed model were inadequate? Matters are complicated as nowadays it has become a standard to acknowledge empirical evidence as evidence for action. That means if data suggest a special relation, researchers usually leave out the next step of research, namely to find a substantial explanation from the context sciences for the special relation that has been *prima facie* found on statistical evidence. That would imply to build a functional (causal) model of interrelations between variables. One example here is the endeavour of the last two decades to find biomarkers that allow for early diagnosis of cancer of various types. Any association or correlation between such a biometric variable and the frequency of cancer has been taken seriously and investigated with millions of measurements and researches but – apart from Herceptin for breast-cancer (Breastcancer.org, n.d.) – with no apparent success. Medical researchers have not built a more refined model of the human metabolism (as in the Herceptin case) to explain the correlation (that has been found) on a context level.

It remains an open issue how to reconcile the standards of science and public implementation of scientific results. The institutional barrier between the stakeholders and the hugely differing interests are only two key elements that cause difficulties. There are further vital elements. There is the difference in information between stakeholders. The available frequentist (more objectivist) information might not be relevant for the single persons. Different criteria applied to the accessible information determine the decision. In the context of the journal copies we have illustrated the minimax principle (minimizing maximum loss) and optimizing expected values in the sense of (4). In general, these criteria lead to different decisions. Furthermore, there is neither universal agreement on such criteria nor a generally accepted scientific argument that one is superior over the others. The criteria fit different purposes, which can be related to tasks that involve risk.

Implications

Language is a highly sensitive issue for all activities in education and the problems are even more relevant in the case of probability and risk. The differing usage of risk has to be addressed in teaching explicitly. The diverse applications may help to enhance what is meant in each case and which connotation the ingredients have so that the results may be evaluated accordingly. The use of the term risk within statistical inference (both in the classical test and in the Bayesian framework) makes issues precise but has to struggle with two main problems:

- i., the logic of statistical inference is complicated and misunderstood, which makes the terms more of a myth than precise.
- ii., the use is different from everyday language and different from use in the context of economy and technical safety considerations. The overlap with everyday meaning and use of risk completes the description of the educational challenge.

To impose the technical jargon and meaning of risk to everyday use requires that the technical and mathematical concepts behind are fully perceived, which cannot be expected from the majority of

students. Can changing the technical concepts be a viable alternative to prepare them better to everyday usage? Would we end up with a different statistics methodology, or new probabilistic and statistical concepts that are more suitable for the purpose of the individual? Or should we approach the problem from a more individualistic perspective from teaching and learning and yet aim to perceive the traditional concepts: increase the awareness of overlap in meaning, purpose, and naming the problems and concepts, and get alert to the wider use of similar wordings and notions in everyday contexts?

Psychological problems to separate between probability and utility have their consequences on proposals to introduce risk and utility considerations from early teaching onwards. Carefully chosen and deliberately discussed examples might help to clarify issues. Further psychological factors seem to be:

- The representation of a situation as win or loss influences decisions.
- The personal status of wealth seems to be vital but is largely unexplored.
- Rather than regarding the probability of the adverse events, people regard the impact of outcomes and possibly focus only on the worst case.
- Low probabilities are badly handled and provoke a shift in focus on impact (maximum loss) rather than evaluating a weighted risk in the sense of (4).
- Impact of various outcomes is reduced to impact in the worst case.
- Cumulative risk is often ignored though it may be relevant.

Constituents of risk are decisions; comparisons of decisions are easier to handle (and impose a partial ordering on various actions) than to understand a value attached to a single decision. How good an action is depends on the alternatives that are feasible. This fact can also be used manipulatively (by pre-set alternatives or reducing alternatives on purpose). We recapitulate some of our main themes.

- a. The approach to probability (APT, FQT, SJT) to evaluate the outcomes (of which the adverse event is but one) regulates the character of calculated risks. However, seemingly more objective information might be useful for an institutional stakeholder but of little help for an individual.
- b. The evaluation of the impact of the outcome may be in terms of money (win or loss), dichotomic (0 for correct decision and 1 for wrong decision), by utility of outcome, or even by idiosyncratic ways in judging specific unwanted events (as, e.g., in the three-door problem).
- c. One-off decisions and repeated situations do require a different logic, which confuses and might provoke a focus on subsidiary “information”.
- d. Criteria used for decisions vary, e.g., for different stakeholders and are often very personal. Statistical information is abstract and often ignored while personal experience is vivid and tends to be overestimated; maximum loss seems more convincing than abstract weighting of impact by probabilities in the sense of (4).

The stakeholders’ position has a definite impact on risk calculations and the decisions to make. There seems to be a clash for the logic of risk as implementation of considerations is different for the stakeholders; recommendations are also interest-driven rather than neutral in a scientific sense and differ also by diverse liabilities within systems (e.g., the health system, medical doctors, the individual patient, his or her relatives). What is rational and good for one stakeholder needs not be rational for another stakeholder, at least the connotation of the used concepts changes as does the information that can sensibly be used. Similar difficulties arise in risk-sharing in technological environments.

There are many topics that seem rewarding to put on the agenda, to ask back, and to find more time to clarify issues. Of course, it depends on the background of the learners whether such additional discussions are helpful for them. The many aspects of a concept may also confuse them. However, at least the teacher has to be aware of the circumstance that people have their private concepts (and which ones they have), that the use in everyday terms is ambiguous and diverse, and that even experts have widely diverging concepts and their own specific terminology. Not the least important element to note – experts also have their own interests and have to be challenged whether their advice can be taken as neutral. Statistics education is well-advised to be careful about the ways to teach issues of risk so that students can develop their own concepts and recognize the influential factors of their decisions as well as decisions that are made on a societal level for them.

Acknowledgements

We thank Ramesh Kapadia for valuable ideas and comments that helped to structure the present paper.

References

- Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Chicago: Markham.
- 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, & A. Rossman (eds.), *Statistics in School Mathematics. Challenges for Teaching and Teacher Education* (pp.71-84). New York: Springer.
- Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Didáctica de la Matemática*, 2, 5-27. Online: www.aiem.es/index.php/aiem/.
- Borovcnik, M., & Kapadia, R. (2011a). Risk in health: more information and more uncertainty. *Proceedings IASE Satellite Conference on "Statistics Education and Outreach"* (6pp.). Voorburg: ISI. Online: iase-web.org/Publications.php.
- Borovcnik, M., & Kapadia, R. (2011b). Determinants of decision-making in risky situations. *Proceedings of the 58th World Statistics Congress* (6pp.). Voorburg: ISI. Online: 2011.isiproceedings.org/papers/950138.pdf.
- Borovcnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E.J. Chernoff, & B. Sriraman (eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education*, v. 7 (pp.7-34). Berlin: Springer.
- Breastcancer.org (n.d.). How Herceptin works. Online: www.breastcancer.org/treatment/targeted_therapies/herceptin/how_works.
- CCOHS (n.d.). Hazard and risk. *Canadian Centre for Occupational Health and Safety*. Online: www.ccohs.ca/oshanswers/hsprograms/hazard_risk.html.
- Dubben, H.-H., & Beck-Bornholdt, H.-P. (2010). *Mit an Sicherheit grenzender Wahrscheinlichkeit. Logisches Denken und Zufall*. (With a probability coming close to certainty. Logic thinking and randomness.) Reinbek: Rowohlt.
- Friedman, M. (1976/1962). *Price theory: a provisional text*. Chicago: Aldine.
- 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.
- Hansson, S. O. (2007). Risk. In E. N. Zalta (ed.), *Stanford Encyclopedia of Science*. Online: plato.stanford.edu/archives/spr2014/entries/risk/.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2), 263-292.
- Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.
- Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? *Journal of the American Statistical Association*, 88, (424), 1242-1249.
- LeRoy, S. F., & Singell, Jr., L. D. (1987). Knight on risk and uncertainty. *Journal of Political Economy*, 95 (2), 394-406.
- Merriam-Webster (n.d.). *Encyclopædia Britannica*. Online: <http://www.merriam-webster.com/>
- Neumann, J. v., & Morgenstern, O. (1953). *The theory of games and economic behavior*. 3rd ed. Princeton: Princeton University Press.
- Rasmussen, N., et al. (1975). Reactor safety study: an assessment of accident risks in U.S. commercial nuclear power plants, *WASH-1400*. Washington, DC: U.S. Nuclear Regulatory Commission.
- Ritholtz, B. (n.d.). Defining risk versus uncertainty. *The big picture*. Online: www.ritholtz.com/blog/.
- Sandblom, G., Varenhorst, E., Rosell, J. Löfman, O., & Carlsson, P. (2011). Randomised prostate cancer screening trial: 20year follow-up. *British Medical Journal*, 342. Online: doi:10.1136/bmj.d1539.
- Spiegelhalter, D. (2014). What can education learn from real-world communication of risk and uncertainty? Invited lecture. *8th British Congress of Mathematics Education (BCME 8)*, Univ. of Nottingham. See also: *Harding Center for Risk Literacy*. Online: <https://www.harding-center.mpg.de/en/health-information/facts-boxes/psa>.
- Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, vol.4(1): *Personelle Wahrscheinlichkeit und Rationale Entscheidung*; (2): *Personelle und statistische Wahrscheinlichkeit* (Problems and results of science and analytic philosophy). Berlin: Springer.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211 (1), 453-458.