Economic Decisions under Uncertainty

by Hans-Werner Sinn

North Holland: Amsterdam, New York and Oxford 1983

Chapter 1: The Object of Choice under Uncertainty

Chapter One

The Object of Choice under Uncertainty

Section A
The Basic Decision-Theoretic Approach

1. *The Ordering of Alternatives*

It is the task of preference theory to indicate general criteria by which men choose, or should choose given their preferences, from a set of mutually exclusive action alternatives $(a_1, a_2, \ldots, a_m)$.

The economic approach[1] to a solution of this task is to search for an evaluation function $R(.)$, attaching to each of the action results $(e_1, e_2, \ldots, e_m)$ a real number with the property[2,3]

$$(1) \qquad R(e_i)\left\{\begin{smallmatrix} > \\ = \\ < \end{smallmatrix}\right\}R(e_j) \Leftrightarrow e_i\left\{\begin{smallmatrix} \succ \\ \sim \\ \prec \end{smallmatrix}\right\}e_j$$

such that the optimal alternative can in principle be found by selecting the highest number:

$$(2) \qquad \max_{a_i} R(e_i).$$

---

[1] It was initially developed by PARETO (1906, p. 176). Cf. also the 'reconsideration' by HICKS and ALLEN (1934).

[2] In this book the symbols $\succ$, $\prec$, $\sim$, and $\Leftrightarrow$ mean respectively 'is better than', 'is worse than', 'is equally good as', and 'if, and only if'. The curved brackets indicate that those enclosed symbols which are at the same height belong together; to read them cross-wise is not permitted.

[3] For all decision problems studied in this book, the function $R(.)$ is taken as given. Since, however, in the intertemporal part IV B, e_i is a time path of result variables, the possibility of changes in preferences *derived* for actual decisions is incorporated; indeed we shall find a very characteristic time dependence of derived preferences. Not all decision-theory approaches imply a given function $R(.)$. The Minimax-Regret Principle (cf. fn. 26 in section B) of NIEHANS (1948) and SAVAGE (1951), for example, has the property that the size of the opportunity set affects $R(.)$ an aspect that MILNOR (1954) was right in criticizing. (My present preference is apple $\succ$ pear $\succ$ sandwich. If my choice is between an apple and a pear, I take the apple. If it is between an apple, a pear, and a sandwich I still take the apple.)

The approach does not provide an answer to the question of which action to choose if there is no unique maximum, that is, if the highest value of $R(.)$ can be achieved in more than one way. In this case the decision can be made by plucking the petals of a daisy. The function $R(.)$ may be called a utility function, but it must be clearly recognized that utility in this context is only defined up to a strictly monotonic transformation of a subjective degree of satisfaction, which means that it is ordinal utility.

Of course, it is a necessary condition of this approach that the opportunity set contains only those alternatives which, by virtue of a preference ordering, can be given a unique value. We ensure this condition by the fundamental

Axiom of Ordering: *The decision maker has a complete weak ordering of all attainable action results.*

It primarily indicates[4]

- that, comparing two arbitrary achievable results, the decision maker is able to make the assessment 'not worse than' ($\gtrsim$) and
- that $e_i \gtrsim e_j$ and $e_j \gtrsim e_k$ imply $e_i \gtrsim e_k$ (transitivity).

It is easy to see that the Axiom of Ordering implies the existence of the preference function $R(e_i)$ although nothing more than weak ordering is required. Obviously

$$(3) \qquad \begin{aligned} R(e_i) > R(e_j) &\Leftrightarrow (e_i \gtrsim e_j \quad \text{and not} \quad e_j \gtrsim e_i), \\ R(e_i) = R(e_j) &\Leftrightarrow (e_i \gtrsim e_j \quad \text{and} \quad e_j \gtrsim e_i). \end{aligned}$$

Instead of a weak ordering of preferences being assumed, a strong ordering could have been postulated at the outset. Probably, however, the decision maker finds it easier to make an assessment of 'not worse

[4] A more complete list of the implications of this axiom can be given as follows. Let X be the Cartesian product attained by multiplying the set of all possible results by itself. Let Y denote the set of all pairs of results the decision maker is able to order by virtue of the relation $\gtrsim$, and let Y' be the converse of Y. Then we have:

$$\begin{aligned} & X \subseteq Y \cup Y' && \text{(completeness)}, \\ \text{in general it is not true that } & \forall e_i, e_j : e_i \gtrsim e_j \Rightarrow e_j \gtrsim e_i && \text{(non-symmetry)}, \\ & e_i \gtrsim e_j \text{ and } e_j \gtrsim e_k \Rightarrow e_i \gtrsim e_k && \text{(transitivity)}. \end{aligned}$$

An implication of completeness is the reflexivity of the relation $\gtrsim$, that is, $e_i \gtrsim e_i$. The set of ordered pairs for which the (reflexive, symmetrical, transitive) equivalence relation $\sim$ holds is $Y \cap Y'$. If $\bar{Y}'$ denotes the complement of Y' in Y then the set of all ordered pairs of results for which the strong preference relation $>$ is valid is $Y \cap Y'$. Since the relation $>$ is irreflexive it is also non-symmetric. Cf., e.g., NACHTKAMP (1969, pp. 66–81) and FISHBURN (1970, pp. 9–15).

than' than 'equally good as' or 'better than'. Moreover, from a behavioristic point of view, the above formulation has the advantage of only utilizing conclusions that can be drawn from observing what people actually choose. By observing a decision it is possible to conclude that the chosen alternative is not worse than the alternatives not chosen, but it is impossible to find out whether the decision maker liked it as much as or more than the others. Some alternative also has to be chosen in the case of indifference[5].

The Axiom of Ordering might appear innocuous and self-evident. However, from both the positive and the normative points of view, it is an idealization. Certainly no one is able to construct a completely consistent ordering of *all* the alternatives available to him in real life[6]. And even if it were possible, people might prefer to do something else from time to time rather than continually investigate preferences and order alternatives. From this, it is evident that a shortcoming of the Axiom of Ordering is its neglect of the *effort of ordering*.

In practical life the effort of ordering implies that the preference function $R(.)$ has a stochastic element[7], so that intransitivities are observable when the action results deviate only a little from one another. This can be explained as follows. Assume the decision maker has some prior information on the function $R(.)$ without knowing its exact value for the various alternatives. Then it is certainly possible that, when comparing e_i with e_j and e_j with e_k, he decides in both cases that one alternative is not worse than the other simply because the advantage he expects to gain from finding the better alternative is not worth the effort of ordering. Although $e_i \sim e_j$ and $e_j \sim e_k$ in this case, we must not conclude that $e_i \sim e_k$ as we could under transitive preferences. The reason is obvious. If the decision maker faces the task of making a decision between e_i and e_k, he is concerned about the advantage to be gained from knowing the true preference ordering between *these* two alternatives. This advantage may exceed that to be gained from knowing the ordering between e_i and e_j as well as that to be gained from knowing the ordering between e_j and e_k. Hence it may induce the decision maker to

5 Thus LITTLE (1950, pp. 14-52) postulated that preference theory deal with acts rather than results. But it surely should be possible to gain insight into the state of mind of a decision maker by asking him about it.

6 AUMANN (1962) therefore has tried to formulate a preference theory without the requirement of completeness.

7 The first of the economists to consider stochastic preferences was GEORGESCU-ROEGEN (1936). In psychology, however, stochastic sensation functions have been discussed since the famous article of THURSTONE (1927).

calculate properly the ordering between e_i and e_k. The result of this calculation is very likely to be $e_i \nsim e_k$[8].

It is certainly desirable to develop an economic preference theory in which the precision of ordering itself is subject to an optimization process. But unfortunately such a theory is not available and cannot be offered here either[9].

2. *Action Results under Uncertainty*

In a world of certainty, the rule $\max_{a_i} R(e_i)$ for finding an optimal action can easily be interpreted. Here e_i is a particular result known with certainty. Its evaluation by use of the function $R(.)$ should not create fundamental problems. In the theory of the household, the result may be a bundle of consumption goods. In the theory of the firm, e_i can often be identified with the level of profit and hence the rule reduces to the well-known aim of profit maximization.

What, however, is the result of an action under uncertainty? Think of an entrepreneur who, despite uncertainty about the future revenue, has to choose one from a set of mutually exclusive investment projects. Could the results we are speaking of be the profit observable *ex post*? This would not make much sense for the decision about the investment project has to be made before knowing how profitable it will be. The basis for a decision, therefore, can only be a result visualized *ex ante*. Such a result has an element of vagueness in it; it can only be represented as a 'random vector' of possible '*ex post* results' or 'subresults':

---

[8] Cf. SCHNEEWEISS (1967a, pp. 35 f. and 81–84) and KRELLE (1957, p. 637; 1961, pp. 112–116; 1968, pp. 21–24). These authors discuss the problem of calculation costs and the possibility of intransitivities being caused by sensation thresholds. The above reasoning unites both aspects since it explains sensation thresholds through calculation costs. For a theoretical explanation of specious intransitivities in terms of automatons see RÖDDING and NACHTKAMP (1978, 1980).

[9] The postulate should not be confused with the aim of the aspiration-level theory, as formulated by SIMON (1957, pp. 241–260), SIEGEL (1957), SAUERMANN and SELTEN (1962), STARBUCK (1963a and b), and others, which includes in the optimization problem the process of information gathering undertaken in order to find the opportunity set. Contrary to first impressions, this theory does not contradict the Axiom of Ordering. This becomes clear if the various possibilities for information gathering are considered as additional actions within the opportunity set. The inclusion of informatin gathering creates a sequential decision problem, but at each point in time there is a given opportunity set of alternatives, one of which has to be chosen. This is completely in accordance with the Axiom of Ordering. To interpret this choice as if the decision maker were merely trying to achieve an aspiration level below the 'true' optimum is a little bit misleading.

(4) $$e_i = (e_{i1}, e_{i2}, \ldots, e_{in}).$$

For this reason TINTNER (1941, p. 301) has called the evaluation function $R(.)$ 'preference functional'. A concrete example of such a random vector is a lottery ticket.

In order to find out what the result vector e_i may be, the decision maker has to take into account the fact that the single *ex post* result depends not only on his own actions, but also on various environmental influences that he can neither manipulate nor perfectly foresee[10]. For the purpose of elucidation, the decision problem may therefore be represented in the form of a case study that can easily be carried out with the aid of the following 'decision' or 'result matrix' originating from VON NEUMANN and MORGENSTERN (1947).

Table 1

action \ class of states of the world	Z_1	…	Z_j	…	Z_n
a_1	e_{11}	…	e_{1j}	…	e_{1n}
⋮	⋮		⋮		⋮
a_i	e_{i1}	…	e_{ij}	…	e_{in}
⋮	⋮		⋮		⋮
a_m	e_{m1}	…	e_{mj}	…	e_{mn}

Here the symbols $(Z_1, \ldots, Z_n)$ denote mutually exclusive classes of states of the world that the decision maker wants to distinguish[11]. The decision maker knows that if he chooses action a_i and the environment dictates class Z_j the subresult e_{ij} will obtain. However, he does not know into which class the true state of the world will fall; this is the particular aspect of the decision problem that emerges under uncertainty.

The matter becomes more complicated if the problem of time is taken into account. In a non-random world, time does not change the nature of the decision problem very much. Action a_i describes a time path of the decision maker's activity that is uniquely associated with a time path of results. Once the optimal activity path is chosen in advance, the individual will stick to it without making new decisions. Things are

[10] Cf. VON NEUMANN and MORGENSTERN (1947, pp. 10 f.).

[11] Note that we have to consider classes of states of the world rather than completely described states themselves. The decision maker will classify the states of the world according to those criteria he is interested in, but not, of course, according to *all* criteria. This distinction is of some importance for the discussion of Bayes's Theorem which occurs below.

different in a stochastic world[12]. Here, even if possible, it would not be wise to maintain a given time path of activity decided upon in the beginning. Suppose the result matrix described above is valid for just one period of time. In the beginning of the period an action has to be taken, and at the end of the period nature reveals the state of the world and the corresponding *ex post* result. It is very likely, then, that the result matrix will, in general, depend on the state of the world obtaining at the end of the previous period. This implies that an action, which yesterday was considered for today, will only by chance coincide with the action which seems optimal today when the result matrix is known. Thus, it is reasonable to postpone making decisions for as long as possible. However, this does not mean that the decision maker will simply abstract from the whole intertemporal problem. Surely, when deciding about today's actions, he has to take into account the fact that its results have an influence on the opportunity set of actions available tomorrow.

These remarks on the intertemporal problem should be enough at this stage. The problem will be taken up again in chapter IV within a somewhat more specific framework. Until then, it is assumed that a choice has to be made a single time only and that, after some time interval during which a revision is impossible, the result becomes known. It will be shown that this assumption, although unrealistic in itself, may serve as a building block for a multiperiod approach.

We are thus back to the decision problem represented by the above decision matrix. The question now is which criteria should be used to evaluate a row of this matrix which describes the '*ex ante* result' e_i of action a_i. The answer is given in two steps. In the present introductory chapter we try to clarify the problem of what information about the classes of states of the world the decision maker needs in order to come to a decision. In chapters II and III we shall try to specify the evaluation function $R(e_i)$.

Section B
Probabilities

1. *Probabilities as Degrees of Confidence*

If the above result matrix is properly specified, then the element e_{ij} really describes all those aspects in which the decision maker is interes-

[12] Occasionally authors avoid the time problem by resort to the tricky, but rather lacking in content, construction in which the Z's are reinterpreted as time paths of states of the world and the a's as strategies in the sense of life philosophies that ought always to be obeyed.

ted of the situation that arises through a coincidence of action a_i with a state of class Z_j. Thus, *ex post*, he does not care under which state of the world a particular result obtained[1]. *Ex ante* however, when the decision is actually made, the class of states of the world under which a particular subresult would occur is of great interest, since it reveals important information on the degree of confidence with which this subresult can be expected.

The degree of confidence, call it g, will be given a more precise meaning later. However, at this stage we want to assume what follows.

Supplement to the Axiom of Ordering: *The result of an action under uncertainty is the random vector*

$$e_i = \begin{pmatrix} g_1 & g_2 & \dots & g_n \\ e_{i1} & e_{i2} & \dots & e_{in} \end{pmatrix},$$

which, in addition to a description of the alternative subresults, contains information on the degrees of confidence with which these subresults may be expected[2].

With this supplement the assumption is made that the way in which a particular subresult is achieved does not affect the evaluation of the vector e_i. This implicitly excludes the situation where gamblers put up a certain stake not only because of their confidence in winning, but also because they have a preference for the way the game is played. The preference for certain, mostly lengthy and complicated, procedures is analogous to the preference of readers of crime novels for situations where there is initially a large number of suspects who are gradually cleared of suspicion, until eventually one of them emerges as being obviously guilty. We neglect this aspect since it does not seem to be significant for the sober economic decisions to be analyzed in this book[3]. The reader is warned, however, to be careful about applying the

---

1 In another context HIRSCHLEIFER (1965, esp. p. 522) argues in favor of a certain type of complementarity between the states of the world and the subresults. This complementarity, however, arises merely from the implicit assumption that the subresults are imperfectly described. The reader interested in this problem is referred to SINN (1980, fn. 7).

2 Note that there is no requirement for either the subresults or the degrees of confidence to be numerically measurable.

3 The neglect of the procedures of the game was the main point to ALLAIS's (1952) criticism of the von Neumann-Morgenstern utility function for which a stronger version of the Axiom of Ordering is required; cf. VON NEUMANN and MORGENSTERN (1947, p. 26).

preference theory developed here to gambling[4].

Nevertheless, to exemplify a problem, reference will often be made to fictitious gambling situations – for example in the following paragraph. But it should not be forgotten that we are always attempting to illustrate serious economic decision making.

To see how to interpret the degrees of confidence or credibility $(g_1, g_2, \ldots, g_n)$ consider the following decision problem. From an urn with black and white balls, one ball is drawn at random after the decision maker has chosen a color. If the chosen color appears, \$ 1000 is paid out, otherwise nothing. The decision maker does not have a preference for a particular color. He knows that the share of black balls is w_1 and the share of white balls w_2. The result matrix for this problem is:

Table 2

probability	w_1	w_2
class of states of the world / actions	Z_1 (black ball is drawn)	Z_2 (white ball is drawn)
a_1 (black is chosen)	$e_{11} = \$ 1000$	$e_{12} = \$ 0$
a_2 (white is chosen)	$e_{21} = \$ 0$	$e_{22} = \$ 1000$

The way in which the degree of confidence concerning the appearance of certain states of the world is to be understood now becomes obvious. Although both actions are alike insofar as the gain is either \$ 1000 or nothing, the decision maker will not usually be indifferent to the color, but will choose the one that characterizes the majority of the balls in the urn. This way, he chooses the action with the greater probability of winning, for the shares of balls, w_1 and w_2, indicate the probabilities with which they will be drawn.

One of the greatest opponents of the idea that probabilities can be interpreted as degrees of confidence was SHACKLE (1952, pp. 5 f., esp. pp. 109–111; 1955, pp. 3–16). He argued that probabilities only have some meaning if the decision maker can repeat his actions, for then probabilities approximate relative frequencies and hence it makes sense

[4] Cf. the critique in ch. III B 1.3 of the approaches by TÖRNQVIST (1945), FRIEDMAN and SAVAGE (1948), FRIEDMAN (1952), and MARKOWITZ (1952). These all infer the evaluation of economic risks from gambling behavior. Cf. also the remarks at the end of section 3.1.2 below.

to select the alternative which promises more frequent success. But if, as in the present context, a decision is made only once, then, he said, probabilities are irrelevant since either one *or* the other case obtains. Instead of probabilities the decision maker is concerned about the *degree of potential surprise* he attaches to the appearance of a particular subresult. Shackle's reasoning, however, is not particularly convincing. If, for example, there are 70 white and 30 black balls in the urn, which subresult would surprise the decision maker more, 'white' or 'black'? Probably the latter, for the probability of its happening is lower. Although Shackle himself saw things differently, the basic concept of potential surprise does not contradict the probability concept. On the contrary, it elucidates the observable fact that probabilities matter even though the decision is not repeated. In the above example, people typically bet on the color for which being not drawn is the more surprising result: this is the color with the higher probability[5].

Although it seems hard to deny this conclusion for the idealized decision situation described above, the question arises of whether this situation has any relevance for real life decision problems. Do we find situations in reality that correspond to the urn experiment?

Ramsey (1931), de Finetti (1937, 1952), and Savage (1952, 1954) see so few parallels with reality that they suggest basing the decision on subjective probabilities[6]. Unlike Shackle, they thus attach numbers to the degrees of confidence $g_1, g_2, \dots, g_n$. These numbers, more or less by chance, obey the rules of probability calculus and can be interpreted without reference to objective probabilities. With the introduction of subjective probabilities, social science decision theory has advanced significantly. A fundamental shortcoming, however, is embodied in its very nature. Probabilities do not exist that are objective in the sense that two equally well informed and equally rational individuals will necessarily agree on them. For this reason the explanatory power of the theory is rather weak. The problem seems to be particularly relevant for Savage (1954, pp. 63–67) who, unlike de Finetti (1937, pp. 16–24) and probably also Ramsey (1931, pp. 187f.), even for urn experiments of the

---

[5] This is the interpretation of Krelle (1957, pp. 648–651) and Nachtkamp (1969, pp. 199 f.). Cf., however, the contributions by Turvey (1949) and Graaf and Baumol (1949), who stress the differences between degree of potential surprise and probability, with particular reference to the non-additivity of Shackle's measure in the case of mutually exclusive events.

[6] For new mathematical developments of this approach see Gottinger (1974). A recent contribution addressing the more fundamental problem of whether, and to what extent, a numerical description of feelings of plausibility is possible is provided by Schneider (1979). Cf. also Stegmüller (1973).

type described above does not want to be pinned down to saying that the rational decision maker calculates with objective probabilities[7].

Thus it is not surprising that LUCE and RAIFFA (1957, pp. 229–302), SCHLAIFER (1959, pp. 2–23; 1969, pp. 106–127, 201–217), PRATT, RAIFFA, and SCHLAIFER (1965, ch. 2 and 3), and RAIFFA (1968, pp. 104–128) make some attempt to rehabilitate objective probabilities by interpreting the subjective probabilities in a narrower sense than Savage does. They start from the idea that the decision maker is able to transform subjective degrees of confidence for the different classes of states of the world into *equivalent objective probabilities* by estimating the relative frequencies with which these classes would occur under a fictitious multiple repetition of the decision situation. This frequency interpretation of subjective probabilities is certainly inadmissible for Savage, but it is compatible with de Finetti's approach, and Ramsey even considered it to be helpful[8].

Transforming subjective degrees of confidence into equivalent objective probabilities has at least three advantages. First, in a very natural way, it removes the difficulty that subjective and objective probabilities may diverge even though the latter are known to the decision maker. Second, *all* decision problems under uncertainty can be reduced to the task of finding a preference functional $R(.)$ for the action result

$$(1) \qquad e_i = \begin{pmatrix} w_1 & w_2 & \dots & w_n \\ e_{i1} & e_{i2} & \dots & e_{in} \end{pmatrix},$$

where e_i is a random vector of subresults, the 'lottery ticket', differing from the one described in the Supplement to the Axiom of Ordering only in that the degrees of confidence have been replaced by equivalent objective probabilities $(w_1, w_2, \dots, w_n)$. Third, operating with these

[7] This is contrary to KEYNES (1921, p. 4): 'The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is *rational* to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.'

[8] Cf. SAVAGE (1954, p. 4), RAMSEY (1931, pp. 158 f. and 187 f.), and DE FINETTI (1937, pp. 18 f.). On the idea of fictitiously repeated decision situations, Ramsey (p. 188) remarks: 'It is this connection between partial belief [in the sense of degree of confidence; the author] and frequency which enables us to use the calculus of frequencies as a calculus of consistent partial belief. And in a sense we may say that the two interpretations are the objective and subjective aspects of the same inner meaning ...'. The compatibility with de Finetti's approach follows from de Finetti's postulate that the decision maker is able to calculate the relative frequency he expects in the future by the use of an objective procedure from his subjective plausibility estimates.

probabilities enables us to use the rich tool box of mathematical probability theory[9].

Despite these advantages, the Luce-Raiffa-Schlaifer approach suffers from a particular deficiency. It does not examine the way in which the rational decision maker transforms degrees of confidence into equivalent objective probabilities. In the next section but one, B 3, we shall deal with this problem. A problem, which needs to be solved beforehand, is to find a more precise meaning for the term 'objective probability', which has, up to now, only been loosely used. This problem is addressed in the following section B 2.

2. *Objective Probability and Real Indeterminateness*

Consider the following questions asked by FISHER (1906, pp. 266–269). An ideal coin is thrown.

i) How large is the objective probability for 'heads' before the throw?
ii) How large is the objective probability for 'heads' when the coin has been thrown, but cannot yet be observed?

Without much hesitation the reader will answer '1/2' to the first question. Perhaps he will also give this answer to the second question, but some confusion becomes apparent after further consideration. At the point in time when the probability is assessed the result of the throw is perfectly determined. Is it not, therefore, the only reasonable answer to say that the probability is either 0 or 1? What sense does it make to interpret subjective probabilities as estimators of objective probabilities according to the Luce-Raiffa-Schlaifer concept if chance can no longer play a role? It seems that we must separate the decision problems of reality into two categories, namely, one category of decisions where conjectures about facts are involved (Is there life on Mars? Do I find oil if I drill here?), and another where the result is really indeterminate (Will it rain tomorrow? Which demand will occur at this price?).

---

[9] This theory is based on only three axioms of KOLMOGOROFF (1933):

i) The probability is a real number in the closed unit interval.
ii) The probability of a certain event is 1 and of an impossible event is 0.
iii) The probability for the appearance of two events out of a set of mutually exclusive events equals the sum of the probabilities of these events.

These questions give rise to the suspicion that there is something wrong with our understanding of an objective probability. In fact, we have confused objective probability with what could be called 'degree of real indeterminateness'.

That the conceptions are not identical[10], becomes particularly evident if we follow an idea developed by LAPLACE (1814, pp. II f.). He argues that the future history of the world is deterministic, since according to the *Principle of Sufficient Reason* each event has a previous cause. 'Today' follows from 'yesterday' according to fixed laws and, according to the same laws, 'tomorrow' will follow from 'today'. Free will, if it existed[11], could be a basis for indeterminateness but neither it nor any alternative basis does exist so that indeterminateness cannot be a feature of the world's future. An unlimited intelligence could foresee the process of events with certainty[12].

In such a world there is no difference in principle between the two questions asked by Fisher. Even if the probability of its being 'heads' is to be assessed *before* the throw, the final result is already determined. It depends on how the person throwing moves his hand and on how the movement of the air influences the fall of the coin, but all this can be

---

[10] A clear distinction between an objective probability and the subjective degree of indeterminateness is not always made. Cf. REICHENBACH (1935, esp. pp. 8–13), KNIGHT (1921, esp. pp. 221 f.), and DE FINETTI (1949, p. 91) who all treat these conceptions as synonymous.

[11] For KNIGHT (1921, p. 221) free will is the genuine cause of indeterminateness in the operation of the world. This, however, is not obvious since, as HEISENBERG (1955, p. 118) correctly remarks, man can do what he wants but he cannot want what he wants.

[12] Laplace seems to have developed his ideas following the apparent success of macroscopic physics in the fields of astronomy and mechanics. At least this is what he draws on to prove his case. In the light of Heisenberg's *Unbestimmtheitsrelationen* (normally translated as *uncertainty principle*, but verbally: *indeterminateness relations*), which tell us that conceptions like place and time, indispensable in a deterministic world, are meaningless in the microcosmos, some doubts about the Laplacian view of the world are appropriate. Instead REICHENBACH's (1925) and HARTWIG's (1956) *Ätialprinzip* seems to be supported. According to this, equal general causes imply equal stochastic distributions of results. The stochastic element in the micro universe typically averages out on the macro level because of the large number of molecules involved. However, the considerable dispersion that has to be taken into account when calculating the explosive power of an atomic bomb (cf. HEISENBERG (1954, p. 135)) and the influence which stochastic mutations have on the process of evolution (cf. MONOD (1971, p. 57 and pp. 141–150) who formulates an antithesis to the Principle of Sufficient Reason) are lucid examples for the effects on the macro world. Despite all this, no one can exclude the possibility that the randomness of micro variables is ultimately a sign of our ignorance about what is really happening. With reference to the philosophical interpretation of Heisenberg's *Unbestimmtheitsrelationen*, Einstein is said to have expressed doubts about whether God throws dice. Whatever the truth may be, we assume a deterministic world in order to demonstrate that the phenomenon of probabilities does not have to be explained by true indeterminateness.

explained from previous causes according to given laws. Thus there is no true indeterminateness, and the only source of subjective indeterminateness is our own ignorance.

But what is the objective probability if everything is predetermined? According to the Luce-Raiffa-Schlaifer concept it is simply that value towards which, in line with the probability theory of VON MISES (1936), the relative frequency of a particular event will stochastically converge when the decision situation is constantly repeated under indistinguishable conditions.

Thus defined, there is an objective probability for both of Fisher's cases, even for the second one. With constant repetitions the result 'heads' will occur in roughly half the throws. It could be argued, rather sophistically, that the repetition need consist only of looking again and again to see whether it is 'heads' or 'tails' without actually repeating the throws. However this would be a violation of the assumption that the experiment has to be repeated under indistinguishable conditions, for the first time the decision maker sees the coin, its position is unknown, while the other times it is known. A realistic example for this case is found in connection with the exploration for oil. The objective probability of finding oil in a particular field is determined by the relative rate of success in *other* fields that are characterized by the same geological data. The probability is not determined by the successful proportion of drills in the field in question.

The example also shows that indistinguishable conditions cannot mean identical conditions, for there are identical conditions only if the drill is always in the same field. 'Indistinguishable' only means 'with identical prior information'. The fact that this prior information is necessarily limited is the reason a stochastic element appears on the scene at all. From experiment to experiment, the uncontrollable influences on the result vary in a way which is deterministic, but which is not systematically connected with the result. So they produce what we call chance[13].

Let us think about what happens if account is taken of some of the previously uncontrolled influences on the result. In this case, a change in the relative frequency of a particular result has to be reckoned with. To illustrate this, consider once more the example of oil exploration and

[13] The reader who believes that chance requires true indeterminateness may open his telephone book, phone each tenth person, and ask him to state his body size. Although perfectly deterministic, the numbers he hears have to be considered statistically as random variables, just as if the names were selected by some randomization machine whose behavior might be really indeterminate. The systematic sample selection methods in statistics make use of this equivalence.

assume additional seismological tests are introduced as a means of extending control over some of the influential factors. Suppose we calculate the relative rate of success on all fields which, with regard to the previous information and the newly introduced tests, can be considered as equal *a priori*. Then, in general, a value different from that found before the new tests will be obtained. Another example of this effect is the problem of calculating the objective probability of loss for an insurance contract. Although the insurance broker's hair will stand on end, let us assume that an insurance company insures all cars for the same premium: since it is reluctant to make the effort of categorizing the vehicles exactly, it practices community rating. The objective loss probability of a particular, arbitrarily chosen, contract can in this case be derived from the relative frequency of losses within the whole stock of contracts, since this stock can be interpreted as a multiple repetition of an experiment of chance with equal prior information[14]. To make the broker's hair lie down again let us now categorize the cars according to their horse power. In this case, the relative loss frequency in a single category approximately measures the objective probability of loss for a single contract belonging to this category and, compared to the case of community rating, we shall now find that this probability takes on a different value.

These considerations imply that there is no such thing as an objective probability in itself. Probabilities can only meaningfully be defined with respect to some prior information. This prior information is the only source of subjective influence on the value of an objective probability. Insofar as two people possess, or consider relevant, different information, for them there are different objective probabilities. In the case of insurance in particular, we cannot exclude the possibility that the objective probability takes on different values from the standpoint of the company and from that of the person insured.

How information influences the probabiliy can easily be understood with the aid of BAYES's Theorem (1763, p. 381, prop. 5)[15]. It may be helpful to interpret this theorem in the light of the insurance example. Let M denote the set of states of the world the company thinks possible, given its prior information. In order to calculate the probability of loss for a particular contract distinguish the states of the world by all

---

[14] A thorough discussion of the conditions under which the insurance case can be interpreted as an experiment of chance can be found in HELTEN (1973, pp. 7–16).

[15] For an experimental approach to the problem of to what extent people are able to estimate probabilities in line with Bayes's theorem, i.e., to what extent they are capable of calculating correctly objective probabilities see EDWARDS and PHILLIPS (1964).

possible relevant criteria[16], e.g., according to the size of the engine, the number of the driver's accident-free years, and the size of the car. Also include as particularly relevant for a description of the state of the world the cases where an accident happens and where it does not. Let $I \subset M$ be the subset of states still possible after the receipt of certain information such as, for example, 'the contract refers to a car of size X and to a driver with Y loss-free years'. Let $Z \subset M$ denote the subset of states of the world that characterize the case of accident and let $\bar{Z}$ be the complementary set. Assume that the corresponding *a priori* probabilities[17] $W(Z)$ and $W(\bar{Z}) = 1 - W(Z)$ are known as well as the conditional probabilities $W(I/Z)$ and $W(I/\bar{Z})$. Then, from

$$(2) \qquad W(I \cap Z) = W(Z)\, W(I/Z) = W(I)\, W(Z/I)$$

we find Bayes's formula

$$(3) \qquad W(Z/I) = \frac{W(I \cap Z)}{W(I)} = \frac{W(Z)\, W(I/Z)}{W(I)} = \frac{W(Z)\, W(I/Z)}{W(Z)\, W(I/Z) + W(\bar{Z})\, W(I/\bar{Z})}.$$

This formula shows how the *a priori* probability $W(Z)$ changes to the *a posteriori* probability $W(Z/I)$ through the receipt of new information. Figure 1 illustrates this. Moreover, it demonstrates the effects of additional information that the company might gain by considering other criteria affecting the probability of loss, such as 'maximum speed' or 'kilometers per year'. They result in the set of possible states of the world being reduced via $I', I'', \ldots, I^{(n)}$ until finally either

$$(4) \qquad I^{(n)} \subseteq Z, \quad \text{so that} \quad W(Z/I) = 1,$$

or

$$(5) \qquad I^{(n)} \cap Z = \emptyset, \quad \text{so that} \quad W(Z/I) = 0,$$

---

[16] A related concept for the estimation of loss probabilities was developed by BAILEY and SIMON (1960) and HELTEN (1974).

[17] Throughout this book $W(.)$ means 'probability of (.)'. In the present context the probabilities can be interpreted in the following way:

$W(Z)$ = share of accidents in the total set of contracts,

$W(\bar{Z})$ = share of accident-free contracts in the total set,

$W(I/Z)$ = share of accidents for contracts of the category 'car size X, number of accident-free years Y' in the total number of accidents, and

$W(I/\bar{Z})$ = share of accident-free contracts of the category 'car size X, number of accident-free years Y' in the total number of accident-free contracts.

is found. The company is now, in effect, the same as the Laplacian intelligence and knows for certain whether or not a particular contract will bring about a loss.

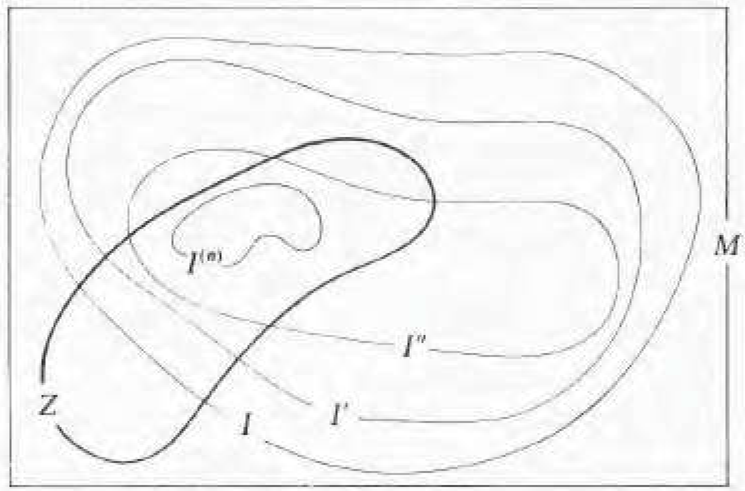

Figure 1

Of course the attempt to increase information will soon face technological and economic barriers that prevent uncertainty from being completely removed. Nevertheless, our thought experiment makes it clear that the objective probability can only be determined for given classification criteria, that is, only after the considered contract is associated with a category of contracts that, while not *identical*, are nevertheless *indistinguishable* with respect to these criteria. *The* unambiguously correct objective probability does not exist. With additional effort it is always possible to obtain more prior information, which leads to a change in the objective loss probability of an arbitrarily chosen contract. This is known as the *Paradox of Homogeneous Grouping* discussed by KNIGHT (1921, pp. 217 f. and p. 224). The perennial question of insurance theory[18], whether a consolidation of non-homogeneous contracts, that is, of those with differing objective probabilities, is possible, is reduced to absurdity by this 'paradox'.

---

[18] Cf. BRAESS (1960, esp. pp. 40 f.). With the interpretation of objective probabilities given above it is not surprising that Lloyd's succeeded in finding a basis for calculating a proper premium for insuring Marlene Dietrich's legs. Although this company probably did not insure many pairs of legs, there were certainly many other risks considered equivalent according to Lloyd's criteria, whatever these may have been.

3. *The Assessment of Equivalent Objective Probabilities*

If an insurance purchaser knows the category into which the company puts him, he does not automatically know his loss probability. Having the information necessary to define an objective probability, therefore, does not imply knowledge of this probability. A probability cannot be known before additional information about a relative frequency of the relevant event in stochastically independent, i.e., not systematically connected, risk situations is available[19].

According to the degree of knowledge of a probability we can distinguish the following categories of decision problems:

– probabilities known with certainty – probability hierarchies known with certainty	risk
– partially known probability hierarchies – completely unknown probabilities	uncertainty

The concept of probability hierarchy used in this list means that alternative probability distributions over the classes of states of the world are considered possible, that for these distributions further alternative probability indications are available, and so on. (A more precise definition will be given in a more appropriate place.) The task will be to reduce the three last cases to the first one[20]. In so doing we shall see that the second case is identical with the first. Both of them are therefore associated with the term 'risk'. In order to demonstrate a fundamental difference, the two latter cases will be called 'uncertainty'.

It would be pleasant if objective probabilities known with certainty were available for real decision problems. But unfortunately this is very rarely the case. Apart from insurance and lotteries there are hardly any practical examples, and even mentioning insurance in this context is not without problems. Why then the never ending discussion in insurance theory about the proper model for loss distribution if this distribution is known with certainty[21,22]?

To gather information on relative frequencies, empirical experience is

---

[19] The conception of 'independence' is used here without judging the philosophical issue of whether independence in the literal sense exists.

[20] Recall that we assume rational behavior. The deficiencies of man in handling probabilities are taken into account here only as a contrast to rational behavior. For a study in various kinds of deficiencies see PHILLIPS (1970) and KAHNEMAN and TVERSKY (1973a and b).

[21] See HELTEN (1973).

[22] Cf., however, SCHNEEWEISS's (1967, p. 271 f.) attempt to defend the case of risk against the contention that it is not very relevant in practice.

not always necessary. KRELLE (1957, p. 638) correctly comments that for many practical problems relative frequencies can be assessed with the aid of thought experiments[23]. This is also the basic concept of the Luce-Raiffa-Schlaifer approach[24], an approach the present book follows in its attempt to reduce all decision problems to the case of risk. Unfortunately we very rarely achieve probabilities by thought experiments for which we would be prepared stand bail.

Thus an analysis of the last three of the above cases is indispensible. We start with the extreme case of completely unknown probabilities and then proceed with the problem of probability hierarchies.

3.1. *Completely Unknown Probabilities*

In the early stages of the development of the theory of uncertainty a number of preference functionals were tailored for this case that went so far as to dispense even with surrogate probabilities, for example, the *Maximin* (or Minimax) *Principle* of WALD (1945; 1950, p. 18) and of VON NEUMANN and MORGENSTERN (1947, p. 101), the *Optimism-Pessimism-Index* of HURWICZ (cited according to MILNOR (1954, p. 50) who refers to an unpublished manuscript) or the *Minimax-Regret-Principle* of NIEHANS (1948) and SAVAGE (1951)[25]. We shall see that the deliberate abstinence from the use of probabilities in these constructions was not really necessary. It *is* possible to find equivalent surrogate probabilities.

3.1.1. The Ellsberg Paradox

Consider a lottery of the kind constructed by ELLSBERG (1961) in order to demonstrate the exact opposite of the above contention.

Out of an urn with white and black balls one ball is drawn randomly. If the decision maker wants to participate in the game he has to pay the price p and to select one of the two colors. If his choice turns out to be

---

[23] GEORGESCU-ROEGEN (1954) seemed to have in mind a similar aspect when he distinguished risk and uncertainty according to whether or not the details of the procedure of a game are known.

[24] Cf. section B 1.

[25] With reference to the result matrix, the preference functional of the Maximin Principle is the minimum of a row, and the preference functional as given by the Optimism-Pessimism Index is a weighted average of a row's minimum and maximum. According to the Minimax-Regret Principle, first each element of the result matrix is replaced by its difference with the maximum of the corresponding column and is multiplied by -1; then the maximum of a row in the so-transformed matrix is taken for the preference functional. A detailed comparison of the criteria is given by MILNOR (1954) and LUCE and RAIFFA (1957, pp. 275–297).

correct he obtains $ 100, otherwise nothing. Table 3 shows the corresponding decision matrix.

Table 3

class of states of the world / actions	black is drawn	white is drawn
black is chosen	$100-p$	$-p$
white is chosen	$-p$	$100-p$
no participation	0	0

Nothing has been said so far about the proportion of white and black balls in the urn. We allow for two alternatives.

(1) *The decision maker knows that the urn contains* $w_1 \cdot 100 = 50$ *white balls and* $w_2 \cdot 100 = 50$ *black balls.*

(2) *The decision maker does not know the relative shares of the two colors.*

The question is, for which of these games he is willing to pay the higher stake.

The typical decision maker answers that his maximum willingness to pay for the first game exceeds that for the second game. This answer demonstrates a choice which is incompatible with the hypothesis that for the second game, the 'uncertainty game', he assigns subjective probabilities to the appearance of the two colors. For, if he does so, he can only have believed one of two things, either that the probabilities are equal ($w_1 = w_2$), or that they are unequal ($w_1 \neq w_2$). In the first case his maximum willingness to pay should be the same for both games. In the second case his maximum willingness to pay for the 'uncertainty game' (2) should be higher than for the 'risk game' (1) since by choosing the right color he has a more than 50% chance of winning.

In the literature there is no unanimous evaluation of this choice known as the *Ellsberg Paradox*. Krelle (1968, pp. 178–184) accepts it as an indication of a particular uncertainty aversion and tries to model this by introducing an *Information Axiom* (p. 181). Roberts (1963)[26] thinks the decision maker misinterprets the decision problem offered to him, and Brewer (1963) and Schneeweiss (1968b) conjecture that the seeming inconsistency between the observed behavior and the use of subjective probabilities arises from the fact that the decision maker is trying to score against the experimenter. It is possible that these interpretations are correct, but perhaps the people interviewed by Ellsberg who where

[26] See also the rejoinder by Ellsberg (1963).

asked to give a quick and intuitive answer were simply being pressed too hard. This interpretation follows from a slight modification of the uncertainty game proposed by RAIFFA (1961).

The decision maker is asked whether, when playing the second game, he prefers a particular color. If, as one would expect[27], his answer is 'no', a third game is suggested:

(3) *A coin is thrown to decide which color to bet on. Then the uncertainty game* (2) *is played.*

When asked to state his maximum willingness to pay for this combined game, the typical decision maker nominates the same amount as for game (2). This seems very reasonable. If he does not have a special preference for one of the two colors then he should be indifferent between choosing the color himself and having it chosen by throwing a coin.

Nevertheless the decision maker has been trapped. It can easily be shown that the third game is virtually indentical with the first game where there is an objective probability of 50% each for the chance of winning and the chance of losing. Consider the following result matrix the head of which shows a tree diagram indicating which classes of states of the world arise from combining the throw of the coin with the draw from the urn.

Table 4

coin throwing decides for	black 1/2		white 1/2	
draw from urn / action	black w_1^*	white w_2^*	black w_1^*	white w_2^*
game(3)	$100-p$	$-p$	$-p$	$100-p$
status quo	0	0	0	0

Let w_1^* and w_2^* denote the unknown probabilities or relative shares of black and white balls. Then, with

$$\tfrac{1}{2}w_1^* + \tfrac{1}{2}w_2^* = \tfrac{1}{2}(w_1^* + w_2^*) = \tfrac{1}{2},$$

we can easily calculate an objective probability of winning and, of course, an equal probability of losing without, and this is surprising, having to know the probabilities w_1^* and w_2^*.

[27] If the decision maker prefers a particular color, we can always repaint the balls.

Comparing the maximum willingness to pay for all three games, we can infer that the decision maker must have made a mistake[28]. With regard to the winning chances, games (1) and (3) are identical. Thus the mistake must either be that games (1) and (2) were considered to be different or that games (2) and (3) were considered to be the same. A clear choice can be made between these alternatives, as long as we accept a well-known axiom for rational behavior under risk.

3.1.2. The Axiom of Independence

In this section Raiffa's trick will be studied in more detail. The analysis will demonstrate that it is irrational to distinguish between risk and uncertainty and will prepare the ground for a rule to be derived in the next section. According to this rule it is possible to achieve equivalent objective probabilities even when starting from a complete ignorance of any probabilities[29].

Consider Table 5. It resembles Table 4, the only difference being that there is the additional game (2) which allows for betting 'white' or 'black'. The reader might be confused at first glance by the fact that for game (2), other than in Table 3, there are result vectors with four elements. However, he may easily convince himself that the throw of a

Table 5

coin throwing decides for		black 1/2		white 1/2	
draw from urn / action		black w_1^*	white w_2^*	black w_1^*	white w_2^*
game (2)	black is chosen	$100-p$	$-p$	$100-p$	$-p$
	white is chosen	$-p$	$100-p$	$-p$	$100-p$
game (3)		$100-p$	$-p$	$-p$	$100-p$
status quo		0	0	0	0

[28] After this explanation the persons asked by Raiffa admitted that they were wrong and revised their evaluations.

[29] To the best of my knowledge Raiffa did not attempt to give a preference-theoretic foundation for his trick. Neither did he extend his 1961 comment on Ellsberg in his later text book (RAIFFA (1968)).

coin is irrelevant and that, whatever color is chosen, there is a chance of winning with the unknown probability w_1^* and a chance of losing with the, also unknown, probability w_2^*.

The essense of the Raiffa trick can now be seen. It is to suggest the equality between game (2) and game (3) by interchanging the last two elements in row 1 or the first two elements in row 2. By doing this Raiffa has implicitly made use of the famous Axiom of Independence that has been frequently discussed in risk theory[30].

The axiom was proposed by MARSCHAK (1950, pp. 120–122, postulate IV), but became popular in a stronger version presented by SAMUELSON (1952a, p. 147). In this book we only consider the strong version[31]; although not really necessary at this stage, this strong version will be needed later.

Axiom of Strong Independence: *Let w denote a (known) objective probability and* e_1, e_2, e_3 *three arbitrary result vectors. (In special cases they may be scalars.) Suppose there is the preference*

$$e_1\{\gtrsim\}e_2,$$

then, for result vectors combined with some arbitrary result e_3,

$$\begin{pmatrix} w & 1-w \\ e_1 & e_3 \end{pmatrix}\{\gtrsim\}\begin{pmatrix} w & 1-w \\ e_2 & e_3 \end{pmatrix},$$

if $0<w\leq 1$.

Referring to concepts from the world of lotteries, we may also express the axiom as follows. If there is a choice between two lotteries, both of which provide the same prize with probability $(1-w)$, but different prizes with probability w, then the ordering of the two lotteries should be the same as that of the two different prizes.

The axiom corresponds to Marschak's weaker version if the indifference symbol $\sim$ only is considered and thus suggests the statement that the value of a lottery is not affected if one of its prizes is replaced by another prize which, though different in kind, is considered to be the same

---

30 In connection with the von Neumann-Morgenstern Index that will be considered below in ch. II C 2.

31 See SAMUELSON (1952a). In SAMUELSON (1952b) a slight modification is introduced. In the version

$$\begin{pmatrix} w & 1-w \\ e_1 & e_1 \end{pmatrix} < \begin{pmatrix} w & 1-w \\ e_2 & e_1 \end{pmatrix} \Leftrightarrow e_1 < e_2$$

the axiom corresponds to the *Sure Thing Axiom* of SAVAGE (1954, p. 73).

from an independent point of view. For this reason, Marschak's postulate is sometimes called the *Substitution Axiom*[32].

The substitution property was used when interchanging the elements within row one and row two in Table 5. Consider the following relations where g_1 and g_2 denote the degrees of confidence in getting, respectively, a black ball and a white ball:

$$\begin{pmatrix} 1/2 & 1/2 \\ e_1 & e_1 \end{pmatrix} = e_1 \equiv \begin{pmatrix} g_1 & g_2 \\ \$\ 100-p & -p \end{pmatrix} \equiv \begin{cases} \text{result of game (2)} \\ \text{when 'black' is chosen} \end{cases}$$

$$\begin{pmatrix} 1/2 & 1/2 \\ e_2 & e_2 \end{pmatrix} = e_2 \equiv \begin{pmatrix} g_1 & g_2 \\ -p & \$\ 100-p \end{pmatrix} \equiv \begin{cases} \text{result of game (2)} \\ \text{when 'white' is chosen} \end{cases}$$

$$\begin{pmatrix} 1/2 & 1/2 \\ e_1 & e_2 \end{pmatrix} \equiv \text{result of game (3)}$$

With regard to the Axiom of Independence, this implies

$$\text{(6)} \qquad \begin{pmatrix} 1/2 & 1/2 \\ e_1 & e_1 \end{pmatrix} \sim \begin{pmatrix} 1/2 & 1/2 \\ e_1 & e_2 \end{pmatrix} \sim \begin{pmatrix} 1/2 & 1/2 \\ e_2 & e_2 \end{pmatrix} \Leftrightarrow e_1 \sim e_2.$$

Thus games (2) and (3) have to be considered as equal if e_1 is not worse than e_2, and e_2 is not worse than e_1, i.e., if the decision maker does not know on which color to bet. In connection with the identity of games (1) and (3) demonstrated above, it follows that it would be a mistake to perceive the uncertainty game (2) as something different from the risk game (1) if the Axiom of Independence is accepted.

To let a coin decide if one does not know which color to choose seems plausible; thus far there is no objection to the Axiom of Independence. There are, however, other implications that at first glance do not seem to be very reasonable. Some of these were taken up by ALLAIS (1952 and 1953).

Suppose you have a maiden aunt, making her last will. She asks you whether you prefer her Rococo sideboard (RSB) or her Colonial sideboard (CSB). Suppose your answer is

$$\text{RSB} \succ \text{CSB}.$$

Unfortunately the aunt intends to leave you her antique clay pitcher (ACP), too, so that you can put it on the sideboard you get. Is your preference in this case automatically

---

[32] For example by ALLAIS (1953, p. 528).

$$(RSB, ACP) > (CSB, ACP)?$$

Obviously not[33].

But this is not what the axiom says. Unlike the example above, it refers to mutually exclusive events. A correct example would be where the aunt offers you a choice between two lottery tickets, the first giving the chance of winning a Colonial sideboard *or* an antique clay pitcher, and the second the chance of winning a Rococo sideboard *or* a clay pitcher. In this case you need not take into account the discomfort you would suffer from seeing the clay pitcher on the Rococo sideboard.

Another criticism[34] cannot so easily be dispensed with. Let us ask whether we could accept the choice

$$\text{(a)} \qquad \begin{pmatrix} 98\% & 2\% \\ \$\,500\text{ mill.} & \$\,0 \end{pmatrix} < \begin{pmatrix} 100\% \\ \$\,100\text{ mill.} \end{pmatrix}$$

and also the other choice

$$\text{(b)} \qquad \begin{pmatrix} 0{,}98\% & 0{,}02\% & 99\% \\ \$\,500\text{ mill.} & \$\,0 & \$\,1 \end{pmatrix} > \begin{pmatrix} 1\% & 99\% \\ \$\,100\text{ mill.} & \$\,1 \end{pmatrix}.$$

Not everyone will decide this way, but many reasonable people do[35]. Define

$$e_1 \equiv \begin{pmatrix} 98\% & 2\% \\ \$\,500\text{ mill.} & \$\,0 \end{pmatrix}, \; e_2 \equiv \begin{pmatrix} 100\% \\ \$\,100\text{ mill.} \end{pmatrix}, \text{ and } e_3 \equiv \begin{pmatrix} 100\% \\ \$\,1 \end{pmatrix}.$$

Then decision (a) is

$$e_1 < e_2$$

and decision (b) is

$$\begin{pmatrix} 1\% & 99\% \\ e_1 & e_3 \end{pmatrix} > \begin{pmatrix} 1\% & 99\% \\ e_2 & e_3 \end{pmatrix}.$$

A comparison reveals a violation of the Axiom of Independence.

33 Cf. ALLAIS (1952, p. 316, footnote). A similar point is made by WOLD (1952); cf. also the directly following discussions with Shackle and Savage and the contribution by SAMUELSON (1952b, pp. 673 f.).

34 ALLAIS (1952, pp. 316 f.; 1953, pp. 529 f.).

35 The majority, but not all, of a group of about 20 students asked by the author. Thus SAMUELSON (1952b, p. 678) does not have to be afraid that he and Savage are the only people in the world able to give consistent answers to Allais's questions.

This example is related to the Ellsberg Paradox. Again it is possible to reveal the inconsistency in the two decisions by presenting them in a slightly different way[36]. Consider first the following problem

$$\text{(c)} \qquad \begin{pmatrix} 1\% & 99\% \\ \begin{pmatrix} 98\% & 2\% \\ \$\ 500 \text{ mill.} & \$\ 0 \end{pmatrix} < \begin{pmatrix} 100\% \\ \$\ 100 \text{ mill.} \end{pmatrix} & \$\ 1 \end{pmatrix}.$$

This formulation means that the decision maker has to participate in a game where, with probability 99%, he can win $ 1 and, with probability 1%, he is given the choice problem (a). It is assumed that his decision would be the same as before. Unless he becomes more optimistic because the 1% chance is realized ('It's my lucky day!') this seems to be a reasonable assumption[37]. Now let us modify (c) by asking the decision maker to announce his decision *before* the outcome of the initial obligatory game is revealed. Except for ALLAIS (1952, pp. 313–330; 1953, p. 538) hardly anyone will come to a different decision. The preference revealed will therefore be

$$\text{(d)} \qquad \begin{pmatrix} 1\% & 99\% \\ \begin{pmatrix} 98\% & 2\% \\ \$\ 500 \text{ mill.} & \$\ 0 \end{pmatrix} & \$\ 1 \end{pmatrix} < \begin{pmatrix} 1\% & 99\% \\ \$\ 100 \text{ mill.} & \$\ 1 \end{pmatrix}.$$

Calculating the probabilities for achieving the alternative possible prizes according to the multiplication rule for independent events, we find that (d) is identical with (b). The fact that the previous decision deviates from the current one reveals the inconsistency. Thus, instead of demonstrating the implausibility of the Axiom of Independence, the example shows that it would have been wise to clarify one's own preference structure with the aid of this axiom before making a decision. As ALLAIS (1953, p. 540) does, we could object to this conclusion on the grounds that problems (b) and (d) are not equivalent since the procedures of the games are different. This, however, is the above-mentioned criticism of the Axiom of Ordering that cannot be accepted for serious economic decision making[38]. The inconsistency revealed in the example can hardly be accounted for by appealing to the pleasure of gambling. Among the people asked by the author whether there is a meaningful difference

[36] The presentation is of the kind chosen by MARKOWITZ (1970, pp. 220–224) for similar examples. Cf. also SAVAGE (1954, p. 103).

[37] Since there are given objective probabilities it is certainly irrational to believe that the probabilities of the second round depend on the result of the first.

[38] Cf. the above remarks to the Axiom of Ordering in section B 1.

between (b) and (d) there was no one who thought that there was[39]. The true explanation of the inconsistency is that presentation (b) hides to some extent the significance of small probabilities while this significance is obvious in (d).

The conclusion to be drawn from Allais's criticism, therefore, is the same as that from the Ellsberg Paradox. Man's capacity to calculate is occasionally strained in decision making under uncertainty. He makes mistakes, he does not behave as his ideal relative modelled in this book does. But he does try to emulate his relative. This is his raison d'être.

3.1.3. A Rehabilitation of the Principle of Insufficient Reason

Equipped with the Axiom of Independence, we are now in a position to generalize the coin-throwing trick. The result of this generalization is the famous *Principle of Insufficient Reason* that dates back to J. Bernoulli (1713, pp. 88 f.) and Laplace (1814, pp. IV and VII)[40]. Applied to our problem, according to this principle the same objective probability has to be attributed to all alternative classes of states of the world as long as the decision maker has no reason to believe that one class is more likely than another[41].

A good example of the value of the Principle of Insufficient Reason is provided by the above model of an urn with *known* content. The likelihood of any one ball being drawn out of an urn seems to be equal to that of any other when all the balls are the same from a manufacturing point of view. Thus we conclude that the relative share of balls of one color indicates the probability of the appearance of this color. That this conclusion is correct can be tested experimentally by repeated drawings.

Is there a different decision problem when it is known which colors are in the urn, but when the shares of these colors are unknown? The difference does not seem to be fundamental for one can well imagine that drawing samples from an infinite number of urns, each of which is known to contain only black balls and white balls, will, in the long run, provide 50% black and 50% white balls. However, many people will feel there *is* a difference. We therefore want to see if there is another way to show that, in the case of complete ignorance of any probability information, it is wise to behave as if there were equal objective probabilities, known with certainty, for all classes of states of the world in the decision problem.

The analysis refers to the following decision problem. There are the classes of states $Z_1, Z_2, \ldots, Z_n$ the objective probabilities of which,

---

[39] Cf. fn. 35.

[40] Keynes (1921, pp. 41 f.) utilized the name *Principle of Indifference*.

[41] The following discussion draws heavily on Sinn (1980).

$w_1^*, w_2^*, \ldots, w_n^*$, $\sum_{j=1}^{n} w_j^* = 1$, are all completely unknown. There is no reason for the decision maker to think that the appearance of one of these classes is more likely than the appearance of any other. The opportunity set consists of the actions $a_1, a_2, \ldots, a_m$ with the result vectors $e_1, e_2, \ldots, e_m$,

$$e_i = \begin{pmatrix} w_1^* & w_2^* & \ldots & w_n^* \\ e_{i1} & e_{i2} & \ldots & e_{in} \end{pmatrix} \quad \forall i = 1, 2, \ldots, m. \tag{7}$$

We thus have a decision problem as illustrated by the matrix of Table 1. The aim is to show that

$$\begin{pmatrix} w_1^* & w_2^* & \ldots & w_n^* \\ e_{i1} & e_{i2} & \ldots & e_{in} \end{pmatrix} \sim \begin{pmatrix} 1/n & 1/n & \ldots & 1/n \\ e_{i1} & e_{i2} & \ldots & e_{in} \end{pmatrix} \quad \forall i = 1, 2, \ldots, m. \tag{8}$$

We first introduce a 'random generator' with states $Z'_1, Z'_2, \ldots, Z'_n$ occuring with objective probabilities $w'_1, w'_2, \ldots, w'_n$, and distinguish the classes of states of an artificial world according to the states of this generator and the classes of states of the real world, such that there are n^2 different classes. The above decision problem can also be demonstrated in this artificial world. This is done with reference to action a_i and result vector e_i in the first row of Table 6. Independently of the states of the generator, subresult e_{ij} occurs if the class Z_j of the states of the real world obtains.

Table 6

random generator		w'_1 → Z'_1	w'_2 → Z'_2	w'_3 → Z'_3		w'_n → Z'_n
real world		w_1^* w_n^* → $Z_1 \ldots Z_n$	w_1^* w_2^* w_n^* → Z_1 $Z_2 \ldots Z_n$	w_1^* w_2^* w_3^* w_n^* → Z_1 Z_2 $Z_3 \ldots Z_n$	$\ldots$	w_1^* w_n^* w_n^* → $Z_1 \ldots Z_{n-1}$ Z_n
a_i	(1)	$e_{i1} \ldots e_{in}$	e_{i1} $e_{i2} \ldots e_{in}$	e_{i1} e_{i2} $e_{i3} \ldots e_{in}$	$\ldots$	$e_{i1} \ldots e_{in-1}$ e_{in}
	(2)	$e_{i1} \ldots e_{in}$	e_{in} $e_{i1} \ldots e_{in-1}$	e_{i1} e_{i2} $e_{i3} \ldots e_{in}$	$\ldots$	$e_{i1} \ldots e_{in-1}$ e_{in}
	(3)	$e_{i1} \ldots e_{in}$	e_{in} $e_{i1} \ldots e_{in-1}$	e_{in-1} e_{in} $e_{i1} \ldots e_{in-2}$	$\ldots$	$e_{i1} \ldots e_{in-1}$ e_{in}
		$\vdots$ $\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$ $\vdots$ $\vdots$ $\vdots$		$\vdots$ $\vdots$ $\vdots$
	(n)	$e_{i1} \ldots e_{in}$	e_{in} $e_{i1} \ldots e_{in-1}$	e_{in-1} e_{in} $e_{i1} \ldots e_{in-2}$	$\ldots$	$e_{i2} \ldots e_{in}$ e_{i1}

Before the other rows are considered, $n-1$ new result vectors

$$(9) \qquad e_i^2 \equiv \begin{pmatrix} w_1^* & w_2^* \dots & w_n^* \\ e_{in} & e_{i1} \dots & e_{in-1} \end{pmatrix},$$

$$e_i^3 \equiv \begin{pmatrix} w_1^* & w_2^* \dots & w_n^* \\ e_{in-1} & e_{in} \dots & e_{in-2} \end{pmatrix},$$

$$\vdots$$

$$e_i^n = \begin{pmatrix} w_1^* & w_2^* \dots & w_n^* \\ e_{i2} & e_{i3} \dots & e_{i1} \end{pmatrix}$$

are defined. These result vectors can be produced from the original vector

$$e_i \equiv e_i^1 = \begin{pmatrix} w_1^* & w_2^* \dots & w_n^* \\ e_{i1} & e_{i2} \dots & e_{in} \end{pmatrix}$$

by moving the subresults step by step to the right and adding to the left side the respective subresult that drops off at the right side with each single step. It is important to note that a pair-wise comparison of any arbitrarily chosen result vectors e_i^j and e_i^k will lead to the judgement that e_i^j is not worse than e_i^k and vice versa. Hence

$$(10) \qquad e_i^k \sim e_i^j \qquad \forall j, k = 1, 2, \dots, n.$$

This equivalence indicates that the decision maker does not care under which class of states of the world a particular subresult occurs. In principle, there are two potential reasons for an interest in the class of states of the world. First, there may be a particular state preference in the sense that, given that a certain subresult occurs, *ex post* the decision maker prefers this subresult to occur under class j rather than under class $j+1$. Second, the decision maker *ex ante* would like a certain subresult to be attached to a particular class because he feels that this class is more plausible or less plausible than others. Both reasons are irrelevant however. The first is excluded by the previous assumption (see the introduction to section B) that the subresults are defined in a way that exhausts all relevant aspects of the situation *ex post*. The second is ruled out by the assumption that the decision maker is completely ignorant of any probabilities of the different classes *ex ante*.

We now proceed by gradually building up the other rows of Table 6, maintaining an equivalence with the first row.

Write the result vector as represented by the first row in the form

$$(11) \qquad e_i^1 = \begin{bmatrix} w_2' & & 1-w_2' & \\ & \left(\dfrac{w_1'}{1-w_2'} \cdot \dfrac{w_3'}{1-w_2'} \cdots \dfrac{w_n'}{1-w_2'}\right) \\ e_i^1 & e_i^1 \quad \cdot \quad e_i^1 \quad \ldots \quad e_i^1 \end{bmatrix}.$$

The advantage of this procedure is that we can make use of the Axiom of Independence and can replace the first element (under w_2') by e_i^2 from (9). A retransformation of (11) then yields the second row of the matrix of Table 6. This second row in turn can be written as

$$(12) \qquad \begin{bmatrix} w_3' & & 1-w_3' & \\ & \left(\dfrac{w_1'}{1-w_3'} \; \dfrac{w_2'}{1-w_3'} \cdot \dfrac{w_4'}{1-w_3'} \cdots \dfrac{w_n'}{1-w_3'}\right) \\ e_i^1 & e_i^1 \quad e_i^2 \quad \cdot \quad e_i^1 \quad \ldots \quad e_i^1 \end{bmatrix},$$

so that the first element can be replaced by e_i^3 from (9). Another retransformation brings about row 3 of the matrix. We proceed in this way and substitute step by step $e_i^4, e_i^5, \ldots, e_i^n$ from (9). The result is row n of Table 6. The use of the Independence Axiom in each step of the transformation procedure ensures that this row and the first row have the same value.

The subresults appearing in row n are the same as those in row 1. For the probabilities of the occurence of these subresults in row n we can easily calculate:

$$(13) \qquad \begin{aligned} W(e_{i1}) &= w_1^* w_1' + w_2^* w_2' + \ldots + w_{n-1}^* w_{n-1}' + w_n^* w_n', \\ W(e_{i2}) &= w_2^* w_1' + w_3^* w_2' + \ldots + w_n^* w_{n-1}' + w_1^* w_n', \\ &\;\vdots \\ W(e_{in}) &= w_n^* w_1' + w_1^* w_2' + \ldots + w_{n-2}^* w_{n-1}' + w_{n-1}^* w_n'. \end{aligned}$$

Note that, so far, assumptions about the sizes of the probabilities $w_1', w_2', \ldots, w_n'$ with which the random generator takes on states $Z_1', Z_2', \ldots, Z_n'$ have not been used. Thus we are free to set

$$w_1' = w_2' = \cdots = w_n' = 1/n$$

for example. This has the advantage that these probabilities can be factored out in (13). The result is:

$$(14)\qquad \begin{aligned} W(e_{i1}) &= 1/n(w_1^* + w_2^* + \ldots + w_n^*) = 1/n,\\ W(e_{i2}) &= 1/n(w_1^* + w_2^* + \ldots + w_n^*) = 1/n,\\ &\vdots\\ W(e_{in}) &= 1/n(w_1^* + w_2^* + \ldots + w_n^*) = 1/n. \end{aligned}$$

Since completely analogous reasoning can be used for all other result vectors $e_1, e_2, \ldots, e_{i-1}, e_{i+1}, \ldots, e_m$ of the original decision problem as given by (7), the equivalence asserted in (8) has been proved.

Thus, the following result which rehabilitates the Principle of Insufficient Reason is achieved. Under complete ignorance of probabilities for the classes of states of the world, the decision maker has to evaluate the rows of his decision matrix
(1) as if each class obtained with the same probability and
(2) as if this probability were an objective value known with certainty.

Attempts to rationalize a similar result were provided by CHERNOFF[42] (1954) and MILNOR (1954) but their axioms are quite technical and intuitively not very appealing, at least according to their critics LUCE and RAIFFA (1957, pp. 286–298, esp. pp. 291 and 296). These attempts have little in common with the one made here, either with regard to the axioms or to the idea of the proof. Moreover, it should be stressed that Chernoff and Milnor assume that the subresults are already transformed into von Neumann–Morgenstern utilities[43]. Later we, too, shall use this utility concept. However, this procedure requires the introduction of a further axiom which is not accepted in lexicographic preference theory and thus should not be used unless necessary[44].

3.1.4. Equivalent Probabilities in Tree Diagrams

In many practical situations the decision problem has a structure resembling that illustrated in Figure 2, that is, the classes of states of the world (Z) are obtained if cases, subcases, subcases of subcases, etc. are distinguished.

An interesting question is which probabilities should be assigned to the classes if the decision maker has no idea at all how plausible the branches of a fork are. According to our previous result, it seems

---

[42] Chernoff's approach is not taken up in the book by CHERNOFF and MOSES (1959).

[43] See in particular CHERNOFF (1954, pp. 422 f.) and MILNOR (1954, p. 49). The 'utilities' in these papers are not only numbers which standardize heterogeneous results as in KRELLE (1968, p. 122; cf. also pp. 144 f.) so that a von Neumann-Morgenstern function can be applied to them, but are also values of this function itself. For example, this is indicated by the fact that Milnor (p. 57) assumes that adding a constant to each element of a column of the result matrix does not change the preference ordering of the rows and also by the fact that Chernoff makes use of the expected-utility concept in definitions 7 and 8 and in postulate 8.

[44] The axiom referred to is the Archimedes or Continuity Axiom. Cf. ch. II C 2.1.

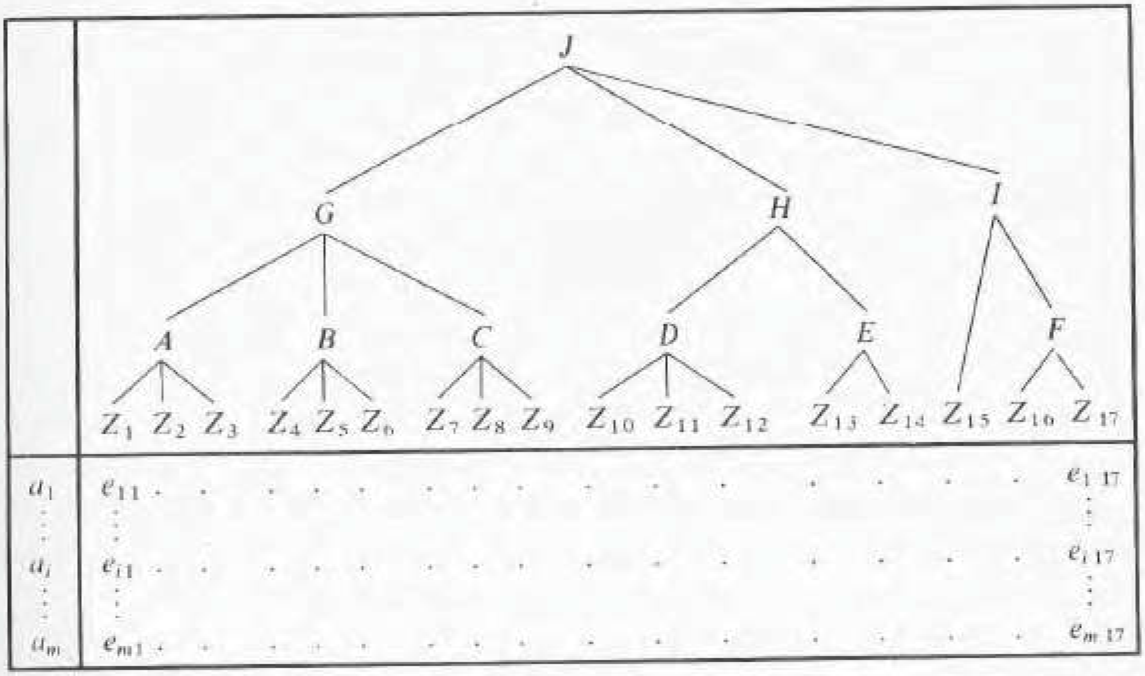

Figure 2

adequate to distribute the probability sum of unity equally among all branches of a fork. Then, according to the multiplication theorem of probabilities, the probability of a certain class of states of the world could easily be calculated by multiplying the probabilities of all the branches from the trunk through to the last small branch defining the class in question. For the example of Figure 2, this method would yield the following probabilities:

no. of class of states of the world	1	2	...	9	10	11	12	13	14	15	16	17
probability	$\frac{1}{27}$	$\frac{1}{27}$	...	$\frac{1}{27}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$

If correct, the important feature of this result is that the Principle of Insufficient Reason yields not only equal probabilities for the classes of states of the world, but also non-uniform probability distributions. The question is, however, whether the result does indeed follow from our axioms.

For simplicity, we refer only to the special case illustrated in Figure 2 and assume that a particular action a_i is chosen, leading to a particular random result vector e_i which is a row in the matrix. We shall consider several subdivisions of this vector that are represented figuratively by the complete set of branches below the forks $A, B, \ldots, J$. The subdivisions are indicated by the letter labelling the corresponding fork.

The demonstration starts with fork A and the corresponding random vector A, consisting of elements e_{i1}, e_{i2}, and e_{i3}. According to the result of the previous section, equal probabilities can be assigned to all branches below fork A without changing the evaluation of vector A. We proceed analogously with forks B, C, D, E, and F and call the result vectors that in this way are assigned objective probabilities $A', B', \ldots, F'$. Without integrating these vectors into the tree diagram at this stage, we now look at fork G and regard it as a random vector consisting of the elements A, B, and C. We then have a problem structured like that in the previous section, for there we did not place any restrictions on what a subresult e_{ij} is. Thus equal probabilities can be assigned to all branches below G (each 1/3). Analogous results can be obtained for the branches below the forks H and I. Now replace elements A, B, and C by A', B', and C' within vector G in a step-wise procedure, referring to the Axiom of Independence:

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ \\ A & B & C \end{bmatrix} \sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ B & C \end{pmatrix} \\ A & \end{bmatrix} \sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ B & C \end{pmatrix} \\ A' & \end{bmatrix}$$

$$(15) \sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ A' & C \end{pmatrix} \\ B & \end{bmatrix} \sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ A' & C \end{pmatrix} \\ B' & \end{bmatrix} \sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ A' & B' \end{pmatrix} \\ C & \end{bmatrix}$$

$$\sim \begin{bmatrix} 1/3 & 2/3 \\ & \begin{pmatrix} 1/2 & 1/2 \\ A' & B' \end{pmatrix} \\ C' & \end{bmatrix} \sim \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ \\ A' & B' & C' \end{bmatrix} \equiv G'.$$

The transformed vector is called G'. Analogously H is transformed to H' and I to I'. The final step is to assign equal probabilities to the branches below J and then to replace G, H, and I by G', H', and I'. Thus, for the example of Figure 2, the Principle of Insufficient Reason has been meaningfully utilized for the assignment of probabilities in tree diagrams. We forego the pure mechanical work of a generalization for arbitrary tree diagrams and state the following.

In case studies for determining the classes of states of the world, whenever no subcase is more plausible than any other, each subcase must be assigned an equivalent objective probability equal to the reciprocal value of the number of subcases. The probability of a certain class is then the product of the probabilities of all cases and subcases that have to be distinguished to define this class.

3.1.5. Criticism of the Principle of Insufficient Reason

Our results are far from being generally accepted. For example, KRELLE (1961, pp. 99 and 106; 1968, pp. 180 f. and 189 f.) refuses to accept the particular aspect reported under point (2) at the end of the last section but one. He does not deny that, in the case of complete ignorance, it is wise to utilize equal subjective probabilities. But he argues that these probabilities should not be treated as if they were objective and known with certainty. Instead, he maintains, the decision maker's preferences may well exhibit a particular kind of 'uncertainty aversion' that cannot be discredited as irrational. On the other hand, with this 'Reduction', and 'Substitution Axioms' Krelle assumes the (weak) Axiom of Independence as we do. Thus, from his point of view, the basic judgement (10) has to be denied. This, however, will hardly be possible without refusing to accept the Axiom of Ordering (also used by KRELLE (1968, pp. 123–125)) and thus rejecting our approach as a whole.

A criticism has also often been made of the *classical* Principle of Insufficient Reason and this needs to be scrutinized to see if it affects our result. A coin is thrown twice. What is the probability that tails comes up both times? If we distinguish the classes of states of the world 'tails, tails' and 'not: (tails, tails)', then the probability sought is 1/2. If, however, we distinguish the classes 'tails, tails', 'tails, heads', 'heads, tails', and 'heads, heads' then the probability is 1/4, a contradiction. Here, the correct solution is obvious, but ascertaining the probability of getting tails at least once can be more confusing. Accordingly, d'Alembert[45], the *enfant terrible* of classical mathematics, argues that if 'heads' comes up with the first throw, a second throw is superfluous. For this reason, the classes of states 'heads, heads', 'tails, heads', and 'tails, tails' should be distinguished, and the probability sought is 2/3 instead of 3/4, the correct probability.

These examples lead us to the problem of which are the classes of states of the world that have to be distinguished in practical decision making, a problem that has already been clearly discussed by VON KRIES (1886, esp. pp. 1–23). Obviously a calculation of objective probabilities according to the Principle of Insufficient Reason demands correctly distinguished classes of states of the world. In the light of classical probability theory, this is a very important problem that unfortunately has never been satisfactorily solved. However, our results are only slightly affected, for we sought subjective probabilities rather than objective ones, although, of course, the former have the form of

[45] According to TODHUNTER (1865, pp. 258 f., art. 464) cited from d'Alembert, Croix ou Pile, Encyclopédie ou Dictionnaire Raisonné ... 1754.

equivalent objective probabilities. In order to make the point quite clear: if d'Alembert does not see any reason why one of his three cases is more plausible than the others, he should indeed assign probabilities of 1/3 to each.

This, however, does not mean that there *is* no reason. Had d'Alembert considered the following tree diagram (Figure 3), he would have found that no one branch is more plausible than the others and thus would have calculated the correct probabilities 1/2, 1/4, and 1/4.

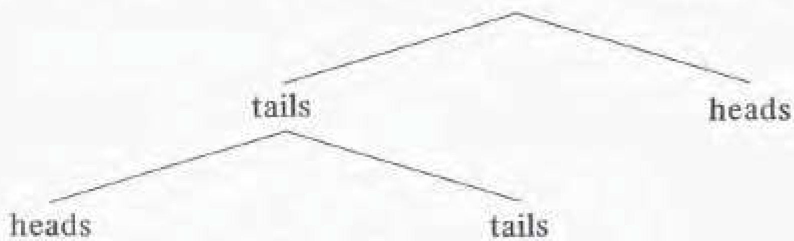

Figure 3

A problem closely related to d'Alembert's was presented by SAVAGE (1954, p. 65). The decision maker can distinguish several possible ways of organizing the classes of states of the world but does not know which organization is the right one. In this case, the Principle of Insufficient Reason seems to fail, for different probabilities can be calculated for a particular event. Consider Savage's example. Two balls are drawn from an urn that is known to contain either two white balls, two black balls, or one white ball and one black ball. If we regard these possibilities as the classes of states of the world, the probability of, for instance, drawing one white and one black ball is 1/3. For Savage, however, it also seems possible to distinguish the classes, 'white, white', 'black, black', 'black, white', and 'white, black', so that the probability in question is 1/2. Fortunately, the problem can be solved. If Savage does not know any reason why one way of organizing the classes is more likely than any other, he may refer to the tree diagram of Figure 4 and assess the probabilities according to the rule developed in the previous section. As a result, he will obtain an equivalent objective probability of 5/12 for drawing one white ball and one black ball.

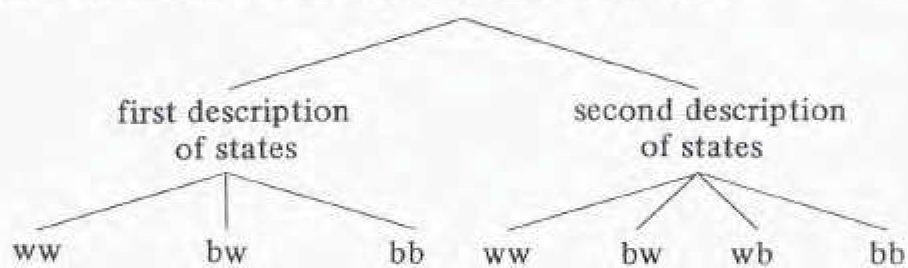

Figure 4

3.2. *Partially Known Probabilities: The Step Theory of Probability*

Until now, we have only considered the cases of probabilities known with certainty and probabilities completely unknown. The reality, however, is between these extreme cases. Typically the decision maker will have more confidence in some variates of the result vector than in others, but will by no means feel sure about his judgement if asked to assess equivalent objective probabilities. He will think instead in terms of alternative probabilities whose correctness he assesses with differing degrees of confidence. Asked whether he is able to transform these degrees of confidence into non-random probabilities he will again shrug his shoulders and so on.

The suitable model for this kind of multi-level uncertainty is the step theory of probability developed by REICHENBACH (1935, pp. 305–322). The basic idea of this theory is used here to transform the imprecise information, the decision maker possesses into equivalent objective probabilities. In the framework of preference theory, our exposition represents a generalization of the approaches by TINTNER (1941) and KRELLE (1968, p. 176) as well as of a short note by ROBERTS (1963, p. 329. fn. 5)[46]. The generalization has two aspects. First, probability hierarchies are considered in Reichenbach's sense. Second, allowance is made for the fact that alternative variates of the probabilities on some level of this hierarchy are known, but that there is no information on the plausibility of these variates themselves. It seems worth noting that the American school of subjectivists, whether we think of Savage or of Luce, Raiffa, and Schlaifer, has avoided this problem by assuming that subjective probabilities can be assessed all at once by questioning people, by asking them to bet or to take part in games, and by other similar methods[47].

To depict multi-level uncertainty we define the following variables:

$i_j, j>1$	current number of the variate of the probability function of step $j-1$
i_1	current number of the class of states of the world
$W^j(i_j), j>1$	non-random probability that the probability of step $j-1$ takes on the variate with current number i_j
$W^1(i_1)$	non-random probability for the occurence of class Z_{i_1}
$W^j_{i_{j+1}}(i_j), j \geq 1$	i_{j+1}th variate of these probabilities if they are themselves random variables

[46] There are also parallels, however, with an approach of SCHNEEWEISS (1964) and with Bayesian statistics where *a priori* distributions of parameters of other distributions and hence probabilities of the second step are utilized. Cf., e.g., HELTEN (1971).

[47] Cf. WINKLER (1967a and b).

$\hat{W}^1(Z_{i_1})$	implicit non-random probability for the occurence of class Z_{i_1}
$\hat{W}^1_k(Z_{i_1})$	kth variate of this probability if it is a random variable
$\tilde{W}(Z_{i_1})$	equivalent objective probability for the occurence of class Z_{i_1}
*	indicates unknown probabilities

With this notation the structure of probability hierarchies can be described. This will first be done under the assumption that on no step are there completely unknown probabilities.

3.2.1. Completely Known Probability Hierarchies

Probabilities of Step One

If all probabilities of the first step are known then there is a given function $W^1(i_1)$, $\sum_{i_1} W^1(i_1)=1$, associating the classes of states of the world $Z_1, Z_2, \ldots$ with a probability. Thus the case of risk prevails.

Probabilities of Step Two

Now the constancy of function $W^1(.)$ is removed. Various variates $W^1_1(.), W^1_2(.), \ldots$ are possible, which themselves occur with probabilities $W^2(1), W^2(2), \ldots$, where $\sum_{i_2} W^2(i_2)=1$. This is the case of known probabilities on the second step that TINTNER (1941) associated with uncertainty as such. Note that the variates $W^1_1(.), W^1_2(.), \ldots$ comprise complete probability distributions over the classes of states of the world and are not defined separately for each class. This construction does not exclude the possibility that, for single classes, there are non-random probabilities of the first step. In this case, the functions $W^1_1(.), W^1_2(.), \ldots$ simply have to take on the same values over these classes.

From the available information it is possible to calculate an implicitly determined probability of the first step for each class of states of the world. This probability is the equivalent objective probability we are looking for. Since, according to the multiplication theorem of probabilities[48], the probability for a coincidence of class Z_{i_1} and variate i_2 of the probability distribution of the first step is $W^1_{i_2}(i_1)\,W^2(i_2)$, summation over all variates of the probability distribution of the first step yields[49]

$$\text{(16)}\qquad \tilde{W}(Z_{i_1}) = \hat{W}^1(Z_{i_1}) = \sum_{i_2} W^2(i_2)\,W^1_{i_2}(i_1).$$

[48] $W(A\cap B)=W(A)W(B/A)$ in the present case is $W\{W^1_{i_2}(.)\cap Z_{i_1}\}=W^2(i_2)W^1_{i_2}(i_1)$.

[49] According to the rule of addition $W(A\cup B)=W(A)+W(B)-W(A\cap B)$ and because the variates of the probability distribution of the first step are disjoint.

Probabilities of Higher Steps

Of course $W^2(.)$, too, does not have to be a given function. We have a problem of uncertainty on the third step if this function can take on the alternative variates $W_1^2(.)$, $W_2^2(.)$, ... which, by virtue of another function, can be associated with probabilities $W^3(1)$, $W^3(2)$, ... The probability for a coincidence of class i_1, variate i_2 of the probability distribution of the first step, and variate i_3 of the probability distribution of the second step is $W_{i_2}^1(i_1)\, W_{i_3}^2(i_2)\, W^3(i_3)$. Thus, summing up over all variates of the probability distributions of the first and second step we find

$$\tilde{W}(Z_{i_1}) = \hat{W}^1(Z_{i_1}) = \sum_{i_3} \sum_{i_2} W^3(i_3)\, W_{i_3}^2(i_2)\, W_{i_2}^1(i_1). \tag{17}$$

A further generalization is evident. If step j is the first one where non-random objective probabilities are available, while there are only probability distributions of probabilities available on lower steps, then

$$\begin{aligned}\tilde{W}(Z_{i_1}) &= \hat{W}^1(Z_{i_1}) \\ &= \sum_{i_j} \sum_{i_{j-1}} \cdots \sum_{i_3} \sum_{i_2} W^j(i_j)\, W_{i_j}^{j-1}(i_{j-1}) \ldots \\ &\quad \ldots W_{i_4}^3(i_3)\, W_{i_3}^2(i_2)\, W_{i_2}^1(i_1).\end{aligned} \tag{18}$$

$\tilde{W}(Z_{i_1})$ is the implicit objective probability of the first step. It is also the equivalent objective probability of the class of states of the world we were seeking.

Criticism

Equations (16)–(18) have in common the characteristic that a probability distribution for the probability of the first step is replaced by a non-random probability that equals the expected value of the distribution:

$$\tilde{W}(Z_{i_1}) = \hat{W}^1(Z_{i_1}) = E[W_{i_2}^1(Z_{i_1})]. \tag{19}$$

Does it make sense to identify this expected value with an equivalent objective probability? Should we not leave some scope for subjective risk evaluations so that if Z_{i_1} brings about a desired situation the optimist will estimate $\hat{W}^1(Z_{i_1})$ as being higher than in (18) and the pessimist will estimate it as being lower? Such an evaluation would be similar to that revealed in Ellsberg's experiments, namely, that people have a particular aversion for unknown probabilities.

As before, however, this aversion cannot stand up to careful examination. The multi-level uncertainty problem described above can be represented by the following urn experiment. From an urn filled with balls of various colors, a random sample of size n_1 is drawn. Out of this sample a subsample of size n_2 is taken, out of this in turn a subsample of size n_3, and so on until finally only a single ball is taken out. Of course it is assumed that $n_1 > n_2 > \cdots > 0$. Does it make a difference for the degree of likelihood of obtaining a ball with a particular color if one ball is drawn directly from the first urn or if it is drawn in the complicated step-wise procedure just described? Except in the case of a particular preference for gambling which was excluded with the Axiom of Ordering, it does not[50]. Thus decision problems with multi-level probabilities that are ultimately known reduce to the case of pure risk.

This solution is certainly very attractive, for it provides an argument for limiting the analysis to the case of risk. But unfortunately this argument is not particularly strong for, contrary to Tintner's contention, crucial aspects of the uncertainty problem have not yet been taken into account.

3.2.2. Partly Known Probability Hierarchies

The task of assessing probabilities on higher steps will soon put too much strain on the decision maker. For example, he might think that alternative variates of the probabilities of the second step are possible, but feel incapable of discriminating between them according to their degree of plausibility.

In such cases the Principle of Insufficient Reason can again be consulted. If the decision maker has no idea which of the alternative variates is more plausible than any other, he has to behave as if all variates were known to occur with an equal objective probability.

This can easily be shown for the general case. Suppose probabilities on step $j+1$ are unknown while the probability function of step j may obtain r alternative variates $W_1^j(.), W_2^j(.), \ldots, W_r^j(.)$. According to (18) the implicit probability of the first step will then have the same number of variates $\hat{W}_1^1(Z_{i_1}), \hat{W}_2^1(Z_{i_1}), \ldots, \hat{W}_r^1(Z_{i_1})$, which, in general, are given by

$$\text{(20)} \qquad \hat{W}_k^1(Z_{i_1}) = W^j(k) \sum_{i_{j-1}} \sum_{i_{j-2}} \cdots \sum_{i_3} \sum_{i_2} W_{i_j}^{j-1}(i_{j-1})\, W_{i_{j-1}}^{j-2}(i_{j-2}) \ldots W_{i_4}^3(i_3)\, W_{i_3}^2(i_2)\, W_{i_2}^1(i_1), \; k = 1, 2, \ldots, r.$$

Thus the problem of not knowing probabilities of step $j+1$ reduces to the problem of unknown probabilities of step two. Assume now that a particular action a_i is chosen. Then the result vector can be written in the form

[50] Cf. the remarks about the Axiom of Ordering in section B 1.

(21) $$e_i = \begin{pmatrix} W^{*j+1}(1) & W^{*j+1}(2) & \dots & W^{*j+1}(r) \\ {}_1e_i & {}_2e_i & \dots & {}_re_i \end{pmatrix}$$

with

$${}_ke_i \equiv \begin{pmatrix} \hat{W}_k^1(Z_1) & \hat{W}_k^1(Z_2) & \dots & \hat{W}_k^1(Z_n) \\ e_{i1} & e_{i2} & \dots & e_{in} \end{pmatrix}, \quad k = 1, 2, \dots, r,$$

where the probabilities of step $j+1$ are marked with a star to indicate complete ignorance. Since this formulation is analogous to (8) we can now directly use the Principle of Insufficient Reason as given above and set

(22) $$W^{*j+1}(k) = \frac{1}{r} \qquad \forall k = 1, 2, \dots, r.$$

Since this result can be achieved for any action a_i, the problem of uncertainty on higher steps has been reduced to an equivalent multi-level risk problem as treated in the previous section. Whenever there is an interruption in the probability hierarchy because probabilities are completely unknown beyond a certain step, it is wise to behave as if there were equal objective probabilities for the variates of the probabilities on the next lowest step.

3.3. *Result*

The fundamental problem of decision making in an uncertain world is to reduce all types of risk and uncertainty to a common base. One extreme type is that of pure risk where there are objective probabilities known with certainty. Another extreme type is characterized by a complete ignorance of any probabilities of the possible action-results. It has been shown that all types can be reduced to the type of pure risk, so that it can serve as a basis for further analysis.

For the extreme case of completely unknown probabilities, a simple rule was derived which rehabilitates the Principle of Insufficient Reason. If the decision maker has no idea at all which probabilities to attach to the classes of states of the world, then he should assign the same probability and should evaluate his action as if these probabilities were objective values known with certainty. This result relies on only two, by no means new, but widely accepted axioms: the Axiom of Ordering and the Axiom of Independence.

In practical decision making under uncertainty, case studies are frequently used. The Principle of Insufficient Reason then has to be employed such that each subcase is assigned an equivalent objective probability equal to the reciprocal value of the number of the subcases. The probability of a particular class of states of the world is the product of the probabilities of all cases and subcases that have to be distinguished to define this class. In this way the Principle of Insufficient Reason yields distributions over the classes of states of the world that are not necessarily uniform.

The decision problems that occur most frequently in real life situations do not seem to be characterized either by complete ignorance of objective probabilities or by firm knowledge of such probabilities. If the decision maker is able to make probability estimates, but does not feel sure about them, probabilities on higher steps have to be considered. Two categories can be distinguished.

(1) If probability distributions of probability distributions of ... are known for the classes of states of the world, then an implicitly given objective probability of the first step can be calculated. Thus the case can be reduced to a decision problem under risk.

(2) If on some step the probability distribution takes on alternative variates for which no probabilities on higher steps are known, then decisions have to be made as if these variates occurred with equal objective probabilities. In this way, the multi-level uncertainty problem, too, can be transformed into an equivalent decision problem under risk.