

Decision Making with Belief Functions: Compatibility and Incompatibility with the Sure-Thing Principle

JEAN-YVES JAFFRAY

Université de Paris VI, LID, Paris, France

PETER WAKKER

University of Leiden (AZL), Department of Medical Decision Making, Leiden, The Netherlands

Abstract

This article studies situations in which information is ambiguous and only part of it can be probabilized. It is shown that the information can be modeled through belief functions if and only if the nonprobabilizable information is subject to the principles of complete ignorance. Next the representability of decisions by belief functions on outcomes is justified by means of a neutrality axiom. The natural weakening of Savage's sure-thing principle to unambiguous events is examined and its implications for decision making are identified.

Key words: belief functions, ambiguity, sure-thing principle, decision making under uncertainty, complete ignorance

1. Introduction

Savage (1954) introduced the “sure-thing principle,” a rationality principle for decision making under uncertainty. It gives a more satisfactory foundation to subjective probabilities than had been obtained before by Ramsey (1931) and de Finetti (1937). It also gives a more profound justification for the “independence axiom”; the latter axiom was used in decision making under risk to characterize expected utility maximization (see von Neumann and Morgenstern, 1944, and Herstein and Milnor, 1953).

In Savage's model, all information is probabilizable. In many situations, however, it is not clear what the probabilities should be. Information may be vague or ambiguous. For that reason, alternative representations are used that do not probabilize all information; examples are Zadeh's (1978) possibility measures and the nonadditive probabilities in Schmeidler (1989). In particular, the belief functions, as introduced by Dempster (1967) and Shafer (1976), have proved useful in artificial intelligence and in robust statistics (see Huber, 1981). They are suited to represent situations where only part of the information can be probabilized and the rest of the information is vague or ambiguous. Examples are imprecise samples and incomplete data bases.

The research of Peter Wakker has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

Given the widespread use of belief functions, it is remarkable that only recently were decision criteria for the above-mentioned type of situations proposed and axiomatized, in Jaffray (1989). He uses as a primitive axiom the independence condition with respect to mixtures of belief functions over the outcomes to generalize expected utility; Jaffray (1992) justifies this condition by means of a lower-probability interpretation of belief functions.

The present article provides a more fundamental foundation for decision criteria based on belief functions. First, preferences and rationality requirements are formulated directly in terms of decisions (“acts”), rather than in terms of belief functions as generated by the acts over the outcomes. Second, by making explicit the underlying knowledge space (see section 2), we can use Pearl’s (1988) probability-of-provability interpretation of belief functions, and we can identify the “neutrality axiom” that underlies Jaffray’s (1989) approach. This is similar to the way in which Savage (1954) provided a more fundamental foundation for expected utility than von Neumann and Morgenstern (1944). Finally, we show that the independence axiom for belief functions corresponds to a natural restriction of Savage’s sure-thing principle, i.e., the restriction to unambiguous conditioning events. This restriction is called the weak sure-thing principle.

In our setup, the decision maker uses a two-stage approach to process the information. The first stage deals with the probabilistic information. A deviation from the Bayesian approach occurs only in the second stage, where the vague information is processed. We assume that the vague information does not contain any meaningful structure. The Bayesian approach will nevertheless invoke probability measures to describe this information, even if this requires a subjective input. The intent of our approach, on the contrary, is to preserve full objectivity; entirely vague information should therefore be processed according to the objective symmetry principles of complete ignorance, as laid down in Arrow and Hurwicz (1972) and Cohen and Jaffray (1980, 1983).

This article is organized as follows. Section 2 formally describes the two-stage modeling of information and gives an example. Section 3 shows that the combination of the probabilizable and nonprobabilizable information naturally leads to belief functions. Sections 2 and 3 do not yet invoke decision-theoretic principles. These are invoked in Section 4 and justify the use of belief functions by means of the “neutrality principle.” The latter shows how to combine the principles of complete ignorance with the basic assumption of decision making under risk, which states that the only relevant aspect of a random variable is its probability distribution.

The axiomatic decision model is studied in section 5. Example 5.1 shows that Savage’s sure-thing principle, characteristic for the Bayesian approach to decision making, and the belief function approach are necessarily incompatible. The experimental data of Cohen and Jaffray (1988) suggest that decision makers do abide by the restriction of the principle to unambiguous events. This weakened sure-thing principle provides a new foundation for Jaffray’s (1989) decision model; this model satisfies linearity with respect to mixtures of belief functions and in this sense generalizes expected utility.

In summary, by using a separation into a stage of probabilizable and a stage of non-probabilizable information, by invoking the principles of complete ignorance, and by

weakening Savage's sure-thing principle in a natural way to probabilizable events, this article obtains a foundation for the use of belief functions in decision making.

2. Probabilizable information and complete ignorance

Several models that deal with situations in which there is both vague and probabilistic information have appeared in the literature. A famous recent example in decision theory is Schmeidler (1989), where vague information is processed in a first stage and probabilizable information in a second. Another example is the classical approach to statistics. Here the vague information concerns the true parameter from a parameter set. In a second stage, given the parameter, the information about the value of the statistic that will be observed is entirely probabilized.

In our approach, the order of stages is reversed as compared to the above models. The probabilistic information is processed in the first stage, and the vague information in the second. Obviously, our model, as well as the models of Schmeidler and of classical statistics, is only applicable in certain situations. It is primarily suited to deal with knowledge representations. Dempster (1967) showed that, in fact, it is a generic model for all belief functions. That is, whenever information can be measured by belief functions, a separation into stages, as in this article, is theoretically possible, although different interpretations are favored by several other authors (see, e.g., Shafer (1990)). The following example illustrates the separation into probabilizable information and complete ignorance.

Example 2.1. At closing time, a TV set retailer has to decide whether or not to serve a last customer. If he does, he will miss the concert he plans to attend, but he is certain to sell one more TV. The retailer's profit depends on the price category, L (ow), M (edium), or H (igh), of the new TV bought by the customer. The prevision of the TV set retailer concerning the type of TV that the customer will buy is based on two pieces of information (see figure 1):

- (i) 60% (or 30%, or 10%, respectively) of the customers own a low (l) (or medium (m), or high (h)) price TV;
- (ii) People, when buying a new TV, either remain in the same price range as in the previous purchase or move to the price range directly above.

The first piece of information is clearly probabilizable. In relation to the second piece of information, we assume that the retailer does not possess any additional information. Nonetheless, the combination of these two pieces of information enables the retailer to make claims such as: 10% of the customers will certainly buy H (those possessing h); 60% will certainly buy L or M (those possessing l); 0% of the customers will certainly buy M . Further, 40% of the customers may possibly buy H (those possessing m or h), 90% may possibly buy M (those possessing l or m), 100% may possibly buy M or H , and 90% of the customers may possibly buy L or M (those possessing l or m), etc. These numbers can also be interpreted as upper or lower bounds for percentages. At least 10%, and at most

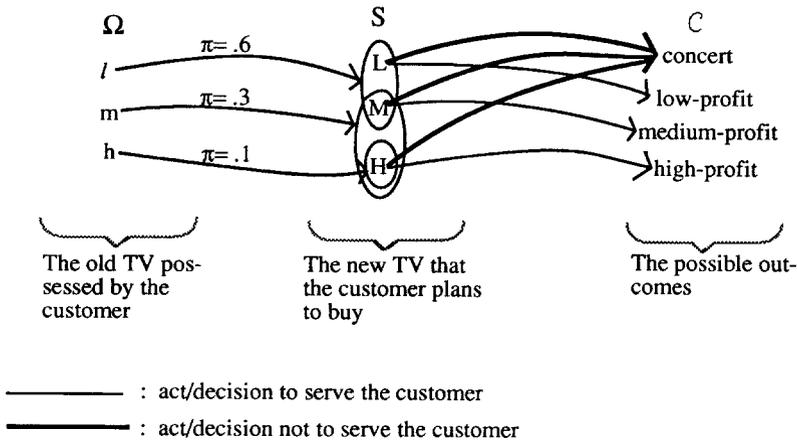


Figure 1. (Example 2.1). We also depict the two decisions; decisions will be discussed in section 5.

40%, of the customers will buy H , and at least 60%, and at most 90%, will buy L or M . We assume that the last customer arrives randomly, so that there is a .1 probability that he/she will certainly buy H , a .4 probability that he/she may possibly buy H ; etc. \square

The information processed in the first stage is modeled through a set Ω , that, for reasons given below, is called the *knowledge space*. Its elements are called *knowledge states*. Exactly one knowledge state is the true one, the others are not true. The decision maker has partial information about which is the true knowledge state. We assume that the information on the set Ω is sufficiently well-structured to be modeled by a probability measure π , in accordance with Savage's (1954) Bayesian approach. In example 2.1, Ω describes the type of TV presently owned by the last customer; with probability .6, this is l , with probability .3, this is m , and with probability .1, this is h . Concerning the remaining uncertainty, given $\omega \in \Omega$, the decision maker has no information at all, and this uncertainty is not probabilizable. The nonprobabilizable information concerns the particular choice of a new TV (superior or the same), given the old TV.

One of the aims of this article is to allow for complete objectivity in the processing of information. This will be an additional motivation for invoking the principles of complete ignorance in the second stage. Complete objectivity is attained if the probability measure on Ω is also objective, for example, if it is derived from statistical data by symmetry considerations, such as in example 2.1. The analysis of this article can also be invoked, however, if that probability measure is not objective.

In decision theory, decisions are modeled as functions from a set S of *states of nature* to an outcome set. One of the states is the true state, the other states are not true, and the decision maker is uncertain about which state is the true state. Subsets of S are called *events*. The outcomes, also called prizes or consequences, describe what will result for the

decision maker. The knowledge space Ω cannot play the role of S because the outcomes of decisions are not solely determined by the ω 's; they partly depend on the nonprobabilizable information.

In example 2.1, outcomes concern the profit that the retailer may obtain and whether or not he attends the concert. We assume that the profit depends on the type of TV sold, hence so does the outcome resulting from the decision to serve the customer. Thus we take $S = \{L, M, H\}$. As the probabilizable information concerns the old TV, we take $\Omega = \{l, m, h\}$. The profit from serving the customer depends only indirectly, and partly, on the old TV. That is to say, the information about the true $\omega \in \Omega$ does not suffice to determine the true state $s \in S$, but leaves uncertainty: all we know is that s describes either the same price range as ω , or that directly above. In other words, all we know is that s belongs to a subset B_ω of S ; $B_l = \{L, M\}$, $B_m = \{M, H\}$, and $B_h = \{H\}$.¹ That no further information is available is expressed by the term "complete ignorance focused on B_ω ," and will be formalized below.

Under risk, i.e., probabilized uncertainty, for each uncertain event a probability has been determined. Under complete ignorance, all that is known about an event is that it is certainly true, certainly not true, or that it is undetermined whether or not the event is true. This is expressed by the truth values: provable, impossible, and undetermined.² We also use the term "possible" for "provable or undetermined." Accordingly, *complete ignorance focused on B_ω* means that B_ω and all of its supersets have truth value provable, \emptyset and all events disjoint from B_ω have truth value impossible, and all remaining events have truth value undetermined.

Suppose, as an illustration, that a given $\omega \in \Omega$ is true, and complete ignorance focused on the event $B_\omega = \{s_1, s_2, s_3\}$ holds. Then the truth values of the events A and A' are the same if:

- (i) $A = \{s_1\}$ and $A' = \{s_2\}$, suggesting a kind of symmetry;
- (ii) $A = \{s_1\}$ and $A' = \{s_2, s_3\}$, showing that cardinality of sets is irrelevant;
- (iii) $A = \{s_1\}$ and $A' = \{s_1, s_2\}$, showing that strict inclusion does not exclude equal truth values.

Let us discuss an example to further illustrate the difference between probabilizable uncertainty and complete ignorance. Consider the set $[0,1]$, endowed with the uniform probability distribution. Usually, this is taken as an instance of total lack of information, related to the principle of insufficient reason as advanced by Laplace (1825). Also it can be obtained through the principle of maximal entropy (see Jaynes, 1968). However, the usual distance function is taken here as a relevant structure; for instance, the statement " $[0, \frac{1}{5}]$ is as large as $[\frac{3}{5}, \frac{4}{5}]$ " is taken as meaningful. This example does not reflect complete ignorance as taken in this article. The information is so well-structured that it can be probabilized. It is not invariant under bijective transformations. Thus, the above-mentioned modeling of insufficient reason for $p \in [0,1]$ has often been criticized for excluding a same modeling of insufficient reason for $p^2 \in [0,1]$. The model of complete ignorance is not open to such criticism; in particular, it is invariant under bijective transformations.

3. Belief functions

Let us now turn to the description of the information concerning the state space S . This combines the probabilizable information concerning Ω and the information given each $\omega \in \Omega$. Throughout the sequel we assume that S is finite. Consider an event $A \subset S$. There are, in view of complete ignorance, only three possible states of information given $\omega \in \Omega$: provable (if A contains B_ω), undetermined (if A intersects B_ω but does not contain B_ω), and impossible (if A does not intersect B_ω). We can partition Ω into three parts, $\Omega_p^A = \{\omega \in \Omega: A \text{ is provable given } \omega\}$, $\Omega_u^A = \{\omega \in \Omega: A \text{ is undetermined given } \omega\}$, and $\Omega_i^A = \{\omega \in \Omega: A \text{ is impossible given } \omega\}$. So the probability that A is provable is $\pi(\Omega_p^A)$, the probability that A is undetermined is $\pi(\Omega_u^A)$, and the probability that A is impossible is $\pi(\Omega_i^A)$. We make the assumption, typical for probability theory, that only these probabilities matter, and not the particular sets $\Omega_p^A, \Omega_u^A, \Omega_i^A$. So the information on S is completely described by mappings $A \mapsto \pi(\Omega_p^A), A \mapsto \pi(\Omega_u^A)$, and $A \mapsto \pi(\Omega_i^A)$. Moreover, it follows from the equalities $\pi(\Omega_i^A) = \pi(\Omega_p^{A^c})$ and $\pi(\Omega_u^A) = 1 - \pi(\Omega_p^A) - \pi(\Omega_i^A)$ that the information on S is in fact completely described by the first mapping alone, henceforth denoted by f .

Note that $f(A) = \pi(\Omega_p^A) = \pi(\{\omega \in \Omega: B_\omega \subset A\})$. If we define the function $\phi: 2^S \rightarrow [0,1]$ by $B \mapsto \pi(\{\omega \in \Omega: B_\omega = B\})$, then

$$f(A) = \sum_{B \subset A} \phi(B). \tag{1}$$

In general, a function $f: 2^S \rightarrow [0,1]$ is called a *belief function on S* if $f(\emptyset) = 0, f(S) = 1$, and f can be written as in (1) for a nonnegative function ϕ on 2^S . We have found above that all the information relevant for the events is contained in the belief function f .

Below we also use the *plausibility function F* , defined by $F(A) = 1 - f(A^c)$. Note that $F(A)$ is the probability that A is possible, i.e., provable or undetermined. Obviously, the plausibility function also contains all information relevant for the events.

For a general belief function f on S , a nonnegative function ϕ as in (1) can always be found and it is uniquely determined (see Rota (1964), Dempster (1967), and Shafer (1976)). To be precise, these references show that $\phi(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} f(B)$. The function ϕ is called the *Möbius transform* of f . Also, an underlying space Ω, π as above can always be constructed, and the above approach to belief functions is general. Thus, by the sole means of the principles of complete ignorance and of probability theory, we have provided a complete justification for belief functions interpreted as probabilities of provability. Alternative justifications for this interpretation have been advanced by Ruspini (1987), Pearl (1988), and Halpern and Fagin (1992).

Example 3.1. (continuation of example 2.1). Table 3.1 describes the belief function, the plausibility function, and the Möbius transform for example 2.1.

Belief functions are *convex*, i.e., $f(A \cup B) + f(A \cap B) \geq f(A) + f(B)$, and even “monotonic of higher orders” (see Shafer, 1976). Convexity implies that the belief function is the infimum of the set of all probability measures that dominate it (see Shapley,

Table 3.1.

	\emptyset	$\{L\}$	$\{M\}$	$\{H\}$	$\{L, M\}$	$\{L, H\}$	$\{M, H\}$	S
$f(B)$	0	0	0	0.1	0.6	0.1	0.4	1
$F(B)$	0	0.6	0.9	0.4	0.9	1	1	1
$\phi(B)$	0	0	0	0.1	0.6	0	0.3	0

1972). This provides an alternative interpretation of belief functions, the so-called lower probability interpretation.

The analysis of this section shows where the Bayesian approach of Savage and the belief function approach of Dempster–Shafer part ways; see also figure 2. It is at the second stage, where events are considered conditional on a knowledge state ω . The Bayesian approach will continue to assign probabilities, conditional given ω , to events, no matter what, leading to a (subjective) probability measure on S .³ The Dempster–Shafer approach can be justified by invoking, conditionally given ω , the principles of complete ignorance, leading to belief functions.

4. A decision-theoretic setup and the neutrality axiom

This section introduces a decision-theoretic approach, in addition to the set Ω , the probability measure π , the set S , the mapping $\omega \mapsto B_\omega$, and the belief function f , as described above (see figure 3). The decision maker must choose between several decision alternatives that we call *acts*. The set of *outcomes* that may result from acts is a finite⁴ set denoted by C . An act d is simply a function from S to C , describing for each state s the outcome that will result if s is true and d is the act chosen by the decision maker.⁵ As the decision maker is uncertain about which element of S is true, he is uncertain about which outcome will result from a chosen act. By \succsim we denote the preference relation of the decision maker over the acts, i.e., we write $d \succsim d'$ if the decision maker considers d at least as good as d' . We assume throughout that \succsim is a *weak order*, i.e., it is *complete* ($d \succsim d'$ or $d' \succsim d$ for each pair of acts d, d') and transitive. We write $>$ for strict preferences, i.e., for the asymmetric part of \succsim , defined by $d > d'$ if $d \succsim d'$ but not $d' \succsim d$, and \sim for indifferences, i.e., for the symmetric part of \succsim , defined by $d \sim d'$ if $d \succsim d'$ and $d' \succsim d$. The preference relation over acts induces a preference relation over outcomes through the constant acts; the latter preference relation is also denoted by \succsim .

Given act d , we denote f_d the mapping that assigns to each subset C of C the value $f(d^{-1}(C))$. Thus $f_d(C)$ is the probability that the event “the outcome of d belongs to C ” is provable, i.e., $f_d(C) = \pi(\{\omega \in \Omega: B_\omega \subset d^{-1}(C)\})$. We denote by ϕ_d the mapping that assigns to each subset G of C the value $\pi(\{\omega: d(B_\omega) = G\}) = \sum_{d(B)=G} \phi(B)$, where ϕ is the Möbius transform of f . Thus $f_d(C) = f(d^{-1}(C)) = \sum_{B \subset d^{-1}(C)} \phi(B) = \sum_{d(B) \subset C} \phi(B) = \sum_{G \subset C} \sum_{d(B)=G} \phi(B) = \sum_{G \subset C} \phi_d(G)$. This shows that f_d is a belief function on C , with Möbius transform ϕ_d . Next we justify the following axiom.

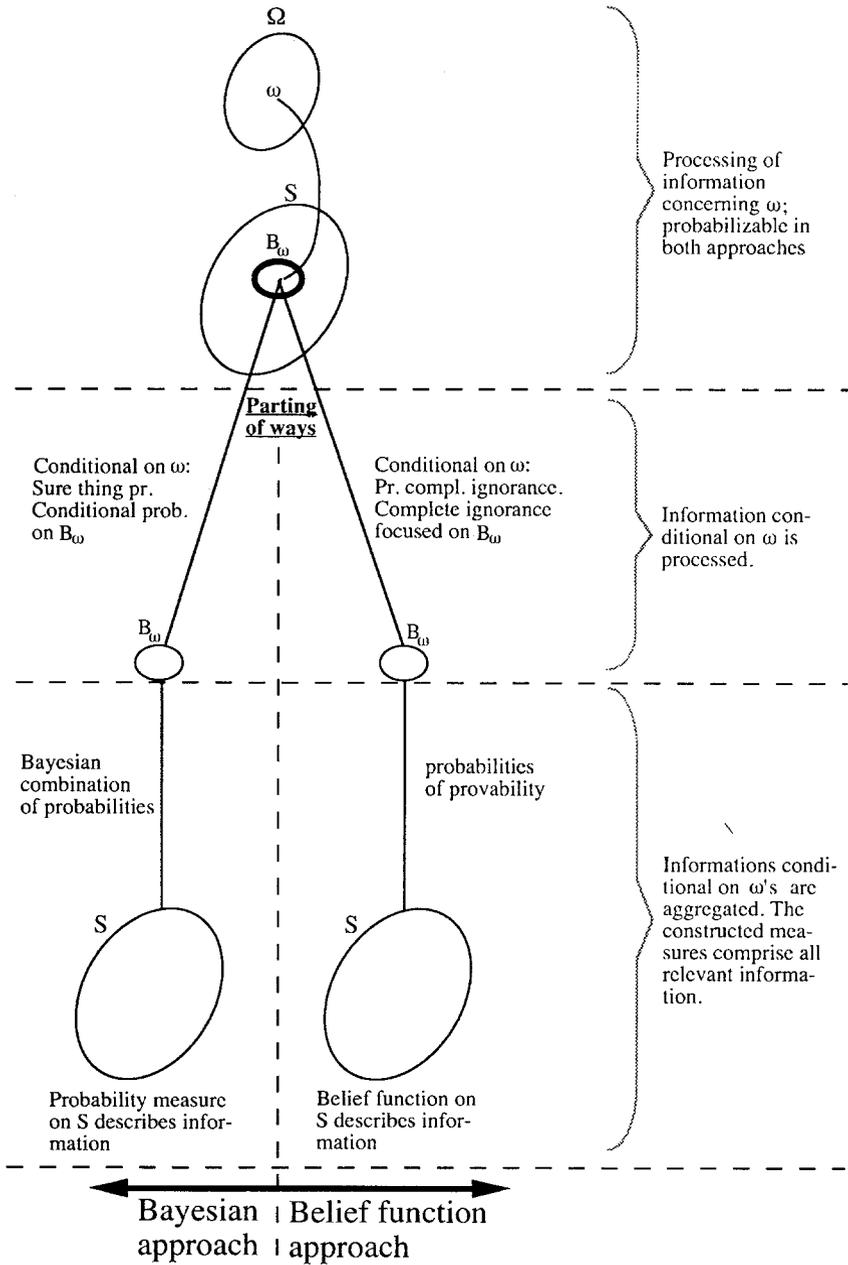


Figure 2. The parting of ways of the Bayesian approach and the belief function approach.

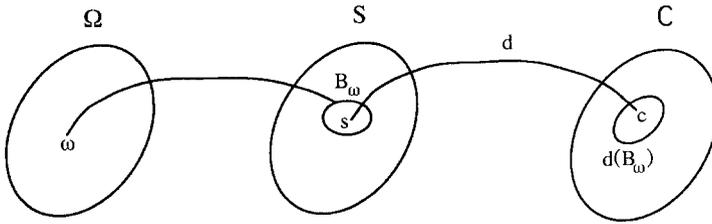


Figure 3. The decision model. An act d maps the state space S to the outcome space C .

Definition 4.1. The *neutrality axiom* is satisfied if acts d and d' be indifferent whenever they generate the same belief function over the outcome space C . □

The neutrality axiom results from the combination of two principles. The first principle, the *principle of choice under complete ignorance*, requires that acts d and d' be indifferent whenever $d(B_\omega) = d'(B_\omega)$ for each ω . For the unconditional one-stage setup where there is only one ω , this principle is axiomatically derived by Arrow and Hurwicz (1972) and Cohen and Jaffray (1980, 1983). The present condition extends the principle to the case of more ω 's. The idea of the condition is as follows. Suppose that $d(B_\omega) = d'(B_\omega)$ for each ω . For each ω and $d(B_\omega) = \{c_1, \dots, c_n\}$, there exist partitions S_1, \dots, S_n and S'_1, \dots, S'_n of B_ω such that $d = c_j$ on each S_j , $d' = c_j$ on each S'_j . All of the S_j and S'_j are either undetermined if $n \geq 2$, or they are provable if $n = 1$, and this is all that is relevant. In particular, the cardinality of the sets S_j and S'_j is not accepted as relevant objective information. The idea of the principle of complete ignorance is that for the evaluation of the act it is irrelevant if, for any ω , the partition S_1, \dots, S_n or S'_1, \dots, S'_n of B_ω is involved. Thus only the range $d(B_\omega)$ given each ω is relevant.

The neutrality principle strengthens the principle of choice under complete ignorance by adding the basic principle of decision making under risk: two random variables that assign the same probability to each element of their range are indifferent. Thus, if for some acts d, d' , and each $G \subset C$, we have $\pi(\{\omega: d(B_\omega) = G\}) = \pi(\{\omega: d'(B_\omega) = G\})$, then $d \sim d'$. The last equality expresses the identity of the Möbius transforms ϕ_d and $\phi_{d'}$, hence of f_d and $f_{d'}$. So, if $f_d = f_{d'}$, then $d \sim d'$, which is exactly what the neutrality axiom requires. *The neutrality axiom will be assumed throughout the rest of the article.*

To discuss critically the setup of this article that builds on the principle of choice under complete ignorance and belief functions, let us consider an example. The deviation from the Bayesian approach occurs only at the second stage, given an $\omega \in \Omega$. Hence, for maximal clarity, only this second stage plays a role in the example. We use the notation e_B for the belief function focused on B , i.e., the function that assigns value 1 to each superset of B and value 0 to all other sets.

Example 4.2. Let $M > m$ be two nonindifferent outcomes. Let $S = \{s_1, s_2, s_3\}$ and suppose that Ω contains only one element ω ; suppose $B_\omega = S$ (so the belief function generated on S is e_S). Define acts d, d' by

$$d(s_1) = M, d(s_2) = m, d(s_3) = m, d'(s_1) = M, d'(s_2) = M, d'(s_3) = m.$$

Then d and d' generate the same belief function on \mathbf{C} , i.e., $e_{\{m, M\}}$. This follows primarily because $\{s_1\}$ and $\{s_1, s_2\}$ have the same truth value, i.e., “undetermined.” By the neutrality axiom, $d \sim d'$. \square

In the above example there is only one ω . Hence, both under d and under d' , each subset $C \subset \mathbf{C}$ has a truth value, and it is the same for d as for d' : provable if C contains both M and m , undetermined if C contains M or m (but not both), and impossible if C contains neither m nor M . By the neutrality axiom, even by the principle of choice under complete ignorance, d' and d should be indifferent. However, d' dominates d , which may be taken as an argument that d' should be strictly preferred to d . The seeming violation of dominance is the price to pay for the complete objectivity that is attained by the principle of choice under complete ignorance underlying the belief function approach. We think that permitting such violations of dominance is not irrational. Note that the violation is only weak in the sense that the dominated act is considered indifferent to the dominating one and is not strictly preferred. Dominance of the type [$d'(s) \geq d(s)$ for all $s \Rightarrow d' \geq d$] is never violated. Dominance violations of the weak type as above are also permitted in the Bayesian approach; then the set where d' dominates d is interpreted as a “null set,” which in the Bayesian frame means it has probability 0. In the belief function approach it can also be said that d' dominates d on a “null set”: the set $\{s_2\}$, when added to the set $\{s_1\}$, does not affect the truth value.

An alternative approach would be to decide, by dominance, that $d' > d$, whereas for the act d'' , defined by $d''(s_1) = m, d''(s_2) = m, d''(s_3) = M$, we would accept both $d' \sim d''$ and $d \sim d''$ in accordance with the principle of choice under complete ignorance. In this case, transitivity of indifference would have to be abandoned. This alternative approach is studied in detail in Cohen and Jaffray (1980).

Next we compare our setup with the axiom that is most characteristic of the Bayesian approach and that is Savage's (1954) most famous invention: the “sure-thing principle.”

Definition 4.3. The *sure-thing principle* is satisfied if, for any two acts that have common outcomes outside a subset $A \subset S$, the preference does not depend on the level of those common outcomes. \square

In other words, the preference between two acts depends only on the restriction of the acts to the event where the acts are different. In the presence of some other natural axioms and some technical axioms, the sure-thing principle characterizes the Bayesian approach (see Savage, 1954). We use example 4.2 to show where our approach violates the sure-thing principle. The acts d and d' are identical outside $\{s_2\} \subset S$; we change their common outcome m on s_3 into the outcome M , so that the following acts result: $\bar{d}(s_1) = M, \bar{d}(s_2) = m, \bar{d}(s_3) = M, \bar{d}'(s_1) = M, \bar{d}'(s_2) = M, \bar{d}'(s_3) = M$. According to the sure-thing principle, the preference between \bar{d} and \bar{d}' should be the same as between d and d' , i.e., \bar{d} and \bar{d}' should be indifferent. Under the belief function approach, however,

it seems natural to strictly prefer \bar{d}' to \bar{d} . The act \bar{d}' dominates \bar{d} in a significant way: under \bar{d}' , the preferred outcome M is certain so that its truth value is “provable”; under \bar{d} , this outcome is not certain and has truth value “undetermined.”

The above development has shown another way to describe the parting of ways of the Bayesian approach and the belief function approach: if, given any ω , one wants to satisfy the sure-thing principle, then the Bayesian approach will result, where one is forced to include information of a subjective nature. If one wants to adhere to objectivity then one will follow the principle of choice under complete ignorance, which in the setup of this article leads to the belief function approach and violates the sure-thing principle. Example 5.1 below further illustrates this, in the richer model of section 5.

Thus, let us turn to a weakening of the sure-thing principle that does not seem reasonable in the belief function approach and that will lead to a specific decision model in the next section. It is illustrated in Figure 4. As a preparation, an event $A \subset S$ is called *ambiguous* if there exists an $\omega \in \Omega$ such that B_ω intersects both A and A^c , so that, for this ω , A has truth value undetermined. Events which are not ambiguous are called *unambiguous*. Note that an event is ambiguous if and only if its complement is. For an unambiguous event A , there exists a subset $A \subset \Omega$ such that $B_\omega \subset A$ for all $\omega \in A$, and $B_\omega \subset A^c$ for all $\omega \in A^c$. Then there is no ambiguity or vagueness about A ; A is true if and only if A is true, and the truth-value undetermined never applies to A . The related probability can be determined as $\pi(A)$, so the event A is in fact probabilizable, with probability $\pi(A) = f(A) = F(A)$. It seems a natural extension of the Bayesian framework to still require the sure-thing principle for the unambiguous events. Violation is permitted only for events that are ambiguous, i.e., that are not probabilizable, and cannot fit into Savage’s framework. Cohen and Jaffray (1988) propose an explanation of the Ellsberg paradox (Ellsberg, 1961) based on the hypothesis that people satisfy the sure-thing principle for unambiguous events.

Thus we formalize:

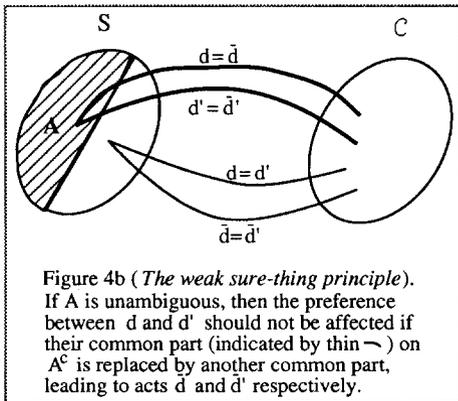
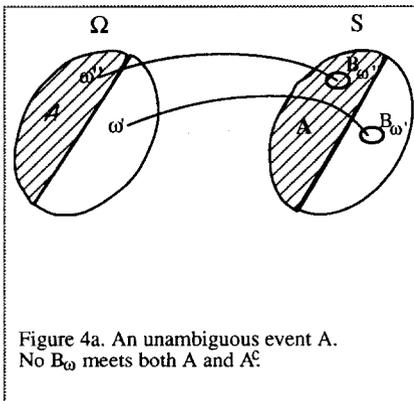


Figure 4. The weak sure-thing principle.

Definition 4.4. The *weak sure-thing principle* is satisfied if, for any two acts that have common outcomes outside an unambiguous event $A \subset S$, the preference does not depend on the level of those common outcomes. □

5. The weak sure-thing principle to characterize a decision theoretic model

In certain cases, the separation of information into two stages, with the set Ω appearing in the first stage, is naturally given. This is the case in example 2.1. In other cases, it is not self-evident how to make such a separation, and it requires creativity on the part of the decision maker. It can then be a major step in the processing of information. Examples and discussions of this point are given in Shafer and Tversky (1985). Also, for different acts, different choices of sets Ω, S can be optimal. We already saw above that, by the neutrality axiom, for decision making the sets Ω, S can be forgotten, as soon as for each act the belief function generated over the outcomes is determined. We base the setup in this section on that observation, and acts are identified with, and usually described as, the generated belief function over \mathbf{C} . Only a few times are sets Ω, S made explicit; these sets may vary from one case to the other. This is entirely in the spirit of probability theory, where distributions of random variables are studied without explicit mention of the underlying probability spaces.

We also discuss the sure-thing principle and its weakened version in this section. These two conditions involve the underlying information structure, i.e., the knowledge space Ω , the space state S , the mapping $\omega \mapsto B_\omega$, and the probability measure π on Ω . Then, of course, it must be assumed that for all four acts involved in the (weak) sure-thing principle, this information structure is the same.

We assume here that the domain of preferences is the set of all belief functions f over the subsets of a finite⁶ set of outcomes \mathbf{C} . The obtained generality of domain is needed for the derivation of the results below.

The following example shows that, except for the trivial case where all outcomes are indifferent, the belief function approach of this section necessarily violates the sure-thing principle, that is, it shows that an example as in section 4 can always be found.

Example 5.1. Take $M, m, S, \Omega, \pi, d, d'$, and the mapping $\omega \mapsto B_\omega$, as in example 4.2. So $d' \sim d$. As $e_{\{M\}} > e_{\{m\}}$, at least one of the indifferences $e_{\{M\}} \sim e_{\{m,M\}} \sim e_{\{m\}}$ must be violated, say the first. Define \bar{d} and \bar{d}' as below definition 4.3, i.e., by changing the common outcome m of d, d' on s_3 into M , so that

$$\bar{d}(s_1) = M, \bar{d}(s_2) = m, \bar{d}(s_3) = M, \bar{d}'(s_1) = M, \bar{d}'(s_2) = M, \bar{d}'(s_3) = M.$$

Then $f_{\bar{d}} = e_{\{m,M\}}$ and $f_{\bar{d}'} = e_{\{M\}}$, and $\bar{d}' + \bar{d}$ follows from $e_{\{m,M\}} + e_{\{M\}}$. This and $d' \sim d$ constitute a violation of the sure-thing principle. □

The above example has shown that the belief function approach and the sure-thing principle are necessarily incompatible. Note that the event $\{s_2\}$ where the acts d and d' , as well as \bar{d} and \bar{d}' , differed, was ambiguous. We shall see below that the belief function approach can be compatible with the sure-thing principle for unambiguous events, i.e., that the weak sure-thing principle can still be respected. To this end, we first introduce mixtures and independence for belief functions.

For two belief functions f, g and $0 \leq \lambda \leq 1$, the mixture $\lambda f + (1 - \lambda)g$ is again a belief function. This follows from the following remark, implied by (1):

$$\text{The Möbius transform of } \lambda f + (1 - \lambda)g \text{ is } \lambda\phi + (1 - \lambda)\psi, \tag{2}$$

where ϕ and ψ are the Möbius transforms of f and g , respectively. Thus nonnegativity of the Möbius transform is preserved under mixing, and mixture of belief functions are again belief functions. So the set of belief functions on C is a convex subset of a linear space (thus obviously a mixture space in the sense of Herstein and Milnor, 1953). In particular, it follows from (1) that

$$f = \sum_{B \subset C} \phi(B)e_B, \tag{3}$$

i.e., each belief function is a convex combination of belief functions focused on subsets of C . In the literature on decision making under risk, mixing, linearity of a representing functional, and independence have been studied when the domain consisted of the probability measures on the set C , leading to expected utility. In our setup this is extended to the larger domain consisting of the belief functions on the set C , leading to a more general theory. We shall obtain a “von Neumann–Morgenstern” representation for preferences, i.e., a representation that is linear with respect to mixtures. First we derive the *independence condition* as used in Herstein and Milnor (1953): for all belief functions f, g, h ,

$$\text{if } f \sim g \text{ then } \frac{1}{2}f + \frac{1}{2}h \sim \frac{1}{2}g + \frac{1}{2}h. \tag{4}$$

Lemma 5.2. The weak sure-thing principle implies independence.

Proof. It suffices to show, for all belief functions f, g, h, \bar{h} , that

$$\frac{1}{2}f + \frac{1}{2}h > \frac{1}{2}g + \frac{1}{2}h \Rightarrow \frac{1}{2}f + \frac{1}{2}\bar{h} > \frac{1}{2}g + \frac{1}{2}\bar{h}, \tag{5}$$

for the following reason: From (5) it follows that: $\frac{1}{2}f + \frac{1}{2}h > \frac{1}{2}g + \frac{1}{2}h \Rightarrow \frac{1}{2}f + \frac{1}{2}f > \frac{1}{2}g + \frac{1}{2}f = \frac{1}{2}f + \frac{1}{2}g > \frac{1}{2}g + \frac{1}{2}g$. Then, by transitivity, $\frac{1}{2}f + \frac{1}{2}h > \frac{1}{2}g + \frac{1}{2}h \Rightarrow f > g$. By contraposition, $f \leq g \Rightarrow \frac{1}{2}f + \frac{1}{2}h \leq \frac{1}{2}g + \frac{1}{2}h$, and twofold application proves the lemma. So we derive (5).

The derivation consists of constructing Ω, S , etc., further acts d, d', \bar{d}, \bar{d}' that generate the belief functions in (5), and an unambiguous event A with probability $\frac{1}{2}$, such that d

and d' coincide outside of A where they generate, “conditionally” (see below), the belief function h . Conditionally on A , they generate the belief functions f and g , respectively. The acts \bar{d}, \bar{d}' are obtained from d, d' by changing them outside of A into acts that generate, conditionally on A^c , the belief function \bar{h} . Then the weak sure-thing principle implies that the preference between d and d' should be the same as between \bar{d}, \bar{d}' , which implies (5). The acts and the generated conditional belief functions have been illustrated in table 5.1.

For the desired construction, take the sets Ω_1, S_1 (S_1 will play the role of the event A below), the probability measure π_1 on Ω_1 , the mapping $\omega_1 \mapsto B_{\omega_1}$ from Ω_1 to 2^{S_1} , and the acts $d_1, d'_1: S_1 \rightarrow \mathbb{C}$, such that $f_{d_1} = f$ and $f_{d'_1} = g$; next take the sets Ω_2, S_2 , the probability measure π_2 on Ω_2 , the mapping $\omega_2 \mapsto B_{\omega_2}$ from Ω_2 to 2^{S_2} , and the acts $d_2, \bar{d}_2: S_2 \rightarrow \mathbb{C}$, such that $f_{d_2} = h$ and $f_{\bar{d}_2} = \bar{h}$. Assume $\Omega_1 \cap \Omega_2 = \emptyset = S_1 \cap S_2$. Finally, define $\Omega = \Omega_1 \cup \Omega_2, S = S_1 \cup S_2, \pi$ is the probability measure on Ω such that $\pi(\Omega_1) = \pi(\Omega_2) = \frac{1}{2}$, and the conditional probabilities given Ω_1 and Ω_2 are identical to π_1 and π_2 , respectively. Let the mapping $\omega \mapsto B_\omega$ from Ω to 2^S coincide with $\omega_1 \mapsto B_{\omega_1}$ on Ω_1 and with $\omega_2 \mapsto B_{\omega_2}$ on Ω_2 . The act d coincides with d_1 on S_1 and with d_2 on S_2 , so $f_d = \frac{1}{2}f + \frac{1}{2}h$. Addition of a prime to d , resulting in d' , means that d_1 on S_1 has been replaced by d'_1 , a bar over the act, resulting in \bar{d} and \bar{d}' , respectively, indicates that d_2 has been replaced by \bar{d}_2 on S_2 . So d' generates the belief function $\frac{1}{2}g + \frac{1}{2}h$, \bar{d} generates the belief function $\frac{1}{2}f + \frac{1}{2}\bar{h}$, and \bar{d}' generates the belief function $\frac{1}{2}g + \frac{1}{2}\bar{h}$. We take $A = S_1$. By the weak sure-thing principle, $d > d' \Rightarrow \bar{d} > \bar{d}'$, implying (5). □

The following continuity condition was used in Herstein and Milnor (1953): \succcurlyeq satisfies *mixture-continuity* if, for each triple of belief functions $f, g, h, \{\lambda \in [0, 1]: \lambda f + (1 - \lambda)g \succcurlyeq h\}$ and $\{\lambda \in [0, 1]: \lambda f + (1 - \lambda)g \preccurlyeq h\}$ are closed. The idea is that a small change in the belief function f_d can be related to a small change in the probabilities $\pi(\{\omega: d(B_\omega) = C\})$, and that such a small change should not induce a large change in preference. By the theorem of Herstein and Milnor (1953), independence and mixture-continuity are equivalent to the existence of a representation V for \succcurlyeq , linear with respect to the mixing of belief functions. Hence, by (3), V is of the form

$$V: f \mapsto \sum_{B \subset \mathbb{C}} \phi(B)V(e_B), \tag{6}$$

where ϕ denotes the Möbius transform of f . We show below that the above form implies the weak sure-thing principle.

Table 5.1.

	d	d'	\bar{d}	\bar{d}'
A	f	g	f	g
A^c	h	h	\bar{h}	\bar{h}

Proposition 5.3. Under mixture-continuity, independence and the weak sure-thing principle are equivalent.

Proof. That the weak sure-thing principle implies independence has already been shown in lemma 5.2. So we derive the weak sure-thing principle, assuming the other conditions of the lemma. Fix Ω, S, π and the map $\omega \mapsto B_\omega$, and let $A \subset S$ be an unambiguous event. Suppose d, d', \bar{d}, \bar{d}' are acts such that d and d' coincide outside of A , and \bar{d}, \bar{d}' are obtained from d, d' by changing the common part outside of A into another common part. If $f(A) = F(A) = 0$, then d and d' generate the same belief function over C , and so do \bar{d} and \bar{d}' , so $d \sim d'$ and $\bar{d} \sim \bar{d}'$, in accordance with the weak sure-thing principle. If $f(A) = F(A) = 1$, then d and \bar{d} generate the same belief function over C and so do d' and \bar{d}' , so $d \sim \bar{d}$ and $d' \sim \bar{d}'$; again, this implies that the preference between d and d' is the same as between \bar{d} and \bar{d}' , in accordance with the weak sure-thing principle.

So we assume henceforth that $0 < f(A) = F(A) < 1$. Now we define Ω_1, Ω_2, S_1 , etc., similar to the proof of lemma 5.2. That is, first we define $S_1 = A, S_2 = A^c$. Because A is unambiguous, Ω can be partitioned into Ω_1 and Ω_2 such that $B_\omega \subset S_1$ for all $\omega \in \Omega_1$ and $B_\omega \subset S_2$ for all $\omega \in \Omega_2$. Thus $\pi(\Omega_1) = f(A) > 0$ and $\pi(\Omega_2) = 1 - F(A) > 0$. We can define the conditional probability distributions π_1 given Ω_1 , and π_2 given Ω_2 . The belief function on C generated by d with respect to π_2 is denoted as h and can be considered to be the belief function generated by d conditionally on Ω_2 , or conditionally on $(S_2 =) A^c$. The belief functions f, g, \bar{h} are defined similarly (see again table 5.1). Writing $\lambda = f(A)$, the belief function on C generated by d is $\lambda f + (1 - \lambda)h$, d' generates $\lambda g + (1 - \lambda)h$, d generates $\lambda f + (1 - \lambda)\bar{h}$, and, finally, \bar{d}' generates $\lambda g + (1 - \lambda)\bar{h}$. Now $\lambda f + (1 - \lambda)h \geq \lambda g + (1 - \lambda)h \Leftrightarrow \lambda f + (1 - \lambda)\bar{h} \geq \lambda g + (1 - \lambda)\bar{h}$ follows from substitution in the representation (6), which is implied by independence, mixture continuity, and the theorem of Herstein and Milnor (1953). It implies that the preference between d and d' is the same as between \bar{d} and \bar{d}' , in accordance with the weak sure-thing principle. □

The above findings lead to the following theorem. It shows that the belief function approach as in (6), while necessarily incompatible with the sure-thing principle in full strength (see example 5.1), is compatible with the weak sure-thing principle and characterizes the latter.

Theorem 5.4. The following two statements are equivalent:

- (i) The preference relation is representable by the form (6).
- (ii) The preference relation satisfies mixture-continuity and the weak sure-thing principle. □

Representations as in (6) have been studied in Jaffray (1989). The above theorem has related that representation to Savage's sure-thing principle and has used this to give a more fundamental foundation than Jaffray's (1989, 1992) foundation. The latter directly

used independence with respect to mixtures. Jaffray (1989) showed that a natural monotonicity condition ($d \succcurlyeq d'$ if $d(s) \succcurlyeq d'(s)$ for all s , for d and d' defined on the same space) implies that V is of the following form:

$$V : f \mapsto \sum_{B \subset C} \phi(B)v(m_B, M_B). \tag{7}$$

Here m_B is the worst outcome of B , M_B is the best outcome of B , and v is a real-valued function such that $v(m'_B, M'_B) \geq v(m_B, M_B)$ whenever $m'_B \succcurlyeq m_B$ and $M'_B \succcurlyeq M_B$; ϕ is the Möbius transform as generated by f on 2^C . The belief function can be interpreted as a description of the (partly vague) information that the decision maker possesses. To interpret the function v , we define a function $u : C \rightarrow IR$ by $u(x) = v(x, x)$, and for each $m \preccurlyeq M$ we define $\alpha(m, M)$ such that $v(m, M) = \alpha(m, M)u(m) + (1 - \alpha(m, M))u(M)$. Substituting this in (6) gives

$$V : f \mapsto \sum_{B \subset C} \phi(B)(\alpha(m_B, M_B)u(m_B) + (1 - \alpha(m_B, M_B))u(M_B)). \tag{8}$$

Here the function u reflects the decision maker's attitude towards outcomes in the context of risk, the function ϕ reflects his state of information and the ambiguity therein, and the function α is an index for his/her like or dislike of the ambiguity. It is well-known that a belief function is additive, i.e., is a probability measure, if and only if its Möbius transform assigns positive values only to sets B that contain exactly one element. For these m_B and M_B coincide and (6), (7), and (8) reduce to expected utility. We conclude that the forms in (6)–(8) are genuine extensions of expected utility to belief functions, i.e., to states of information where not all uncertainty is probabilizable.

Acknowledgements

This article received helpful comments by Michèle Cohen and Alain Chateauneuf.

Notes

1. Note that the sets B_ω need not be disjoint. The often-studied case where the B_ω 's form a partition of S is a special case of the general setup of this article.
2. We choose these terms, more customary in the literature on logic than in decision theory, to stay close to Pearl (1988).
3. In a full Bayesian elaboration, S is partitioned into $S \times \Omega$, and $P(A) = \sum_{\omega \in \Omega} P(A \times \{\omega\})$.
4. The results of this article can, without any difficulty, be extended to the case where C is infinite, due to the fact that, S being finite, decisions generate on 2^C "generalized" belief functions of the form (1). We prefer, however, not to discuss belief functions on infinite sets, hence assume that C is finite.
5. As pointed out above (note 3), in a full Bayesian elaboration, states s would be split up into $\{(s, \omega) : \omega \in \Omega\}$. For assigning outcomes to acts, all elements from $\{(s, \omega) : \omega \in \Omega\}$ are equivalent. Hence they are "lumped together," and the state space is reduced accordingly. The idea of explaining nonadditivity of measures by a lumping of states has been advanced in Gilboa and Schmeidler (1992).
6. The generalization to the set of all "simple" belief functions over the subsets of an infinite set C is straightforward, but, again, for simplicity, we prefer not to discuss belief functions on infinite sets.

References

- Arrow, K.J., and L. Hurwicz. (1972). "An Optimality Criterion for Decision Making under Ignorance," In C.F. Carter and J.L. Ford, *Uncertainty and Expectations in Economics*. Oxford: Basil Blackwell and Mott Ltd.
- Cohen, M., and J.Y. Jaffray. (1980). "Rational Behavior under Complete Ignorance," *Econometrica* 48, 1281-1299.
- Cohen, M., and J.Y. Jaffray. (1983). "Approximations of Rational Criteria under Complete Ignorance," *Theory and Decision* 15, 121-150.
- Cohen, M., and J.Y. Jaffray. (1988). "Is Savage's Independence Axiom a Universal Rationality Principle?," *Behavioral Science* 33, 38-47.
- de Finetti, B. (1937). "La Prévision: Ses Lois Logiques, ses Sources Subjectives," *Annales de l'Institut Henri Poincaré* 7, 1-68. Translated into English by H.E. Kyburg as "Foresight: Its logical Laws, its Subjective Sources." In H.E. Kyburg and H.E. Smokler (eds.), *Studies in Subjective Probability* (1964). New York: Wiley, pp. 53-118; 2nd ed. (1980). New York: Krieger.
- Dempster, A.P. (1967). "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics* 38, 325-339.
- Ellsberg, D. (1961). "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* 75, 643-669.
- Gilboa, I., and D. Schmeidler. (1992). "Additive Representations of Non-Additive Measures and the Choquet Integral," mimeo.
- Halpern, J.Y., and R. Fagin. (1992). "Two Views of Belief: Belief as Generalized Probability and Belief as Evidence," *Artificial Intelligence* 54, 275-317.
- Herstein, I.N., and J. Milnor. (1953). "An Axiomatic Approach to Measurable Utility," *Econometrica* 21, 291-297.
- Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.
- Jaffray, J.Y. (1989). "Linear Utility Theory for Belief Functions," *Operations Research Letters* 8, 107-112.
- Jaffray, J.Y. (1992). "Linear Utility Theory for Belief Functions: A Discussion." In Proceedings of FUR-IV Conference, Budapest (1988). Dordrecht: Kluwer Academic Publishers.
- Jaynes, E.T. (1968). "Prior Probability," *IEEE Transactions on System Sciences and Cybernetics* 4, 227-241.
- Laplace, P.S. de. (1825). *Essai Philosophique sur les Probabilités*, 5th ed. Paris. (Originally published in 1814). Translated into English as *A Philosophical Essay on Probabilities* (1951). New York: Dover.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
- Ramsey, F.P. (1931). "Truth and Probability." In *The foundations of Mathematics and other Logical Essays*. London: Routledge and Kegan Paul, pp. 156-198. Reprinted in H.E. Kyburg and H.E. Smokler (eds.), *Studies in Subjective Probability* (1964). New York: Wiley, pp. 61-92.
- Rota, G.C. (1964). "Theory of Möbius Functions," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2, 340-368.
- Ruspini, E.H. (1987). "The Logical Foundations of Evidential Reasoning." Revised as Technical Note 408, SRI International, Menlo Park, CA.
- Savage, L.J. (1954). *The Foundations of Statistics*. New York: Wiley; 2nd ed. (1972). New York: Dover.
- Schmeidler, D. (1989). "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, 571-587.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.
- Shafer, G. (1990). "Perspectives on the Theory and Practice of Belief Functions," *International Journal of Approximate Reasoning* 4, 323-362.
- Shafer, G., and A. Tversky. (1985). "Languages and Designs for Probability Judgment," *Cognitive Science* 9, 309-339.
- von Neumann, J., and O. Morgenstern. (1944, 1947, 1953). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Zadeh, L.A. (1978). "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems* 1, 3-28.