International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. 8, No. 3 (2000) 271–284

DEMPSTER BELIEF FUNCTIONS ARE BASED ON THE PRINCIPLE OF COMPLETE IGNORANCE

PETER P. WAKKER
Medical Decision Making Unit, Leiden University Medical Center, P.O. Box 9600, Leiden 2300 RC, The Netherlands

Received 29 October 1999
Revised 24 March 2000

This paper shows that a "principle of complete ignorance" plays a central role in decisions based on Dempster belief functions. Such belief functions occur when, in a first stage, a random message is received and then, in a second stage, a true state of nature obtains. The uncertainty about the random message in the first stage is assumed to be probabilized, in agreement with the Bayesian principles. For the uncertainty in the second stage no probabilities are given. The Bayesian and belief function approaches part ways in the processing of the uncertainty in the second stage. The Bayesian approach requires that this uncertainty also be probabilized, which may require a resort to subjective information. Belief functions follow the principle of complete ignorance in the second stage, which permits strict adherence to objective inputs.

Keywords: Belief functions; Complete ignorance; Bayesianism; Nonadditive measures; Ambiguity.

1. Introduction

Belief functions are widely used as an index of belief, alternative to Bayesian additive probabilities. Up to now, there has only been a limited number of studies linking belief functions to decision making.[1–9] This paper considers decision making for belief functions generated in the two-stage manner of Dempster (1967).[10] In the second, final stage, one element of a set of states of nature will obtain, the *true state of nature*. Prior to that, in a first stage a *random message* is received, designating a subset of the state space that will contain the true state of nature. The uncertainty of the first stage, regarding the random message to be received, is probabilizable. That is, it can be expressed in terms of probabilities, in agreement with the Bayesian principles. No probabilities are given for the uncertainty at the second stage, regarding the true state of nature conditional on the received random message. *Throughout this paper we assume this two-stage resolution of uncertainty.* All claims on belief functions are restricted to this setup.

The Bayesian approach requires the use of probabilities in all circumstances, hence also at the second stage, which may necessitate invoking subjective information. Here the belief function approach deviates from Bayesianism. Belief functions can be explained by a resort, in the second stage, to the "principle of complete ignorance" instead of Bayesianism. The principle describes a method for objective decision making in situations where there is a total absence of information.

The analysis of this paper is easiest to state when the probabilities at the first stage are objective and given beforehand. Hence we assume such objective probabilities. This assumption is not essential. The arguments can also be applied when the first-stage probabilities are subjective, e.g., when they are derived from decisions. We further assume that the random messages are observable. This assumption is essential. It must be possible to distinguish between the receipt of different random messages, with a conditioning on each of them being meaningful. Hence the claims of this paper are restricted to Dempster belief functions and need not apply to other belief functions, e.g., in the sense of Shafer (1976).[11]

This paper is based on the first part of Jaffray and Wakker (1993, up to Definition 4.3),[12] where it was already shown informally that the principle of complete ignorance underlies belief functions in a two-stage model of uncertainty. The present text elaborates and formalizes the argument and clarifies the motivation. It generalizes the result to completely general preference relations on general domains that need not satisfy transitivity or completeness.

2. Decision under Uncertainty

$S = \{s_1,\dots,s_n\}$ denotes a finite *state* space. Exactly one state is true, the others are not true, and it is unknown which state is the true one. Subsets of S are *events*. A function f defined on the power set of S is a *belief function* if there exists a function φ from 2^S to [0,1] such that, for each $E \subset S$,

$$f(E) = \sum_{E' \subset E} \varphi(E') \text{ and } f(\varnothing) = 0, f(S) = 1.$$

It is well-known that φ, called the *Möbius inverse* of f, is uniquely determined for each belief function f.[10,11] In the following sections, the sets E' will be random messages and $\varphi(E')$ will be the probability of receiving E' in the first stage. The formula suggests a crude way for assigning Möbius weights. All subsets E' of E are incorporated in the summation, all other sets are completely excluded.

Without further background, the empirical meaning and implications of belief functions are not clear. This paper will give empirical meaning to belief functions by linking them to decision making. We assume the classical model for decision under uncertainty, i.e. besides the state space $\{s_1,\dots,s_n\}$ there is an *outcome* set $\mathcal{C}$, and *acts* map states to outcomes. For an act d, d(s) is the outcome resulting if act d is carried out and s is the true state of nature. Because there is uncertainty about which state of nature is the true one there is uncertainty about what outcome will result from act d.

The empirical primitive in decision theory is a binary relation $\succcurlyeq$, the *preference relation*, over acts. $d \succcurlyeq d'$ means that a decision maker is willing to choose d from {d, d'}. That is, in this paper preference is nothing other than binary choice. *Strict preference* $d \succ d'$ means that $d \succcurlyeq d'$ but not $d' \succcurlyeq d$.

A natural condition for preferences is transitivity. The condition is, however, not needed in the formal analysis. Let me emphasize that we neither need to assume completeness of preference, i.e. it is permitted that between some pairs of acts no choice or preference is observed. The set of acts considered can also be any arbitrary subset of the set of all functions from S to $\mathcal{C}$. In this respect, the approach of this paper is extremely flexible. I will assume *reflexivity* ($d \succcurlyeq d$ for all acts d). Further assumptions on preferences will be formulated later.

The preference conditions presented hereafter are formulated in terms of an equivalence relation ~: Two acts d and d' are *equivalent*, denoted $d \sim d'$, if d' can be substituted for d in each preference. That is, $d \succcurlyeq f$ if and only if $d' \succcurlyeq f$, and $f \succcurlyeq d$ if and only if $f \succcurlyeq d'$. Because of reflexivity, $d \succcurlyeq d'$ and $d' \succcurlyeq d$ whenever d and d' are equivalent. Under common assumptions on preferences, such as weak ordering, equivalence is the symmetric part of preference.

A decision foundation for belief functions is given by Ghirardato, however using a framework different from the classical decision model as just described.[6] Ghirardato assumes that acts are correspondences, assigning sets of outcomes rather than one outcome, to states. In that context he imposes the Savage[13] axioms implying a probability measure on the state space. Then every act generates a probability distribution over subsets of outcomes, thus a belief function over outcomes.

The present paper uses the classical decision model. It will be argued that decisions can be based on (Dempster) belief functions if and only if the so-called principle of complete ignorance is reasonable. Decisions being based on belief functions means that two acts are equivalent whenever they generate the same belief function over outcomes. In other words, for preferences between acts, only the belief functions generated by the acts need to be known. This condition is the analog, for belief functions, of the so-called basic principle of decision under risk. The latter applies when probabilities are given and assumes that acts are equivalent whenever they generate the same probability distribution over the outcomes. It is usually assumed implicitly in decision under risk by describing acts in terms of probability distributions over outcomes without stating the random mechanism generating the outcomes.

The principle of complete ignorance (PCI) is not the only one underlying decisions based on Dempster belief functions. I think, however, that it is the critical one and that the other principles are relatively unobjectionable. Hence Section 3 will discuss the PCI in detail, not yet considering the two-stage model of Dempster. Whereas weak monotonicity principles can always be respected, a strict monotonicity condition has to be abandoned under the PCI. The two-stage model of Dempster is introduced in Section 4, where the main result is also presented. Section 5 concludes.

3. The Principle of Complete Ignorance

This section considers total absence of information, i.e. situations where not any information is available about the states $s_1,\ldots,s_n$.[1] The principles of total absence of information also apply if decisions must be based on objective information but only subjective information is available, a common case in group decisions. In this section, we assume that the outcome set $\mathcal{C}$ contains outcomes \$1 and \$0, with the former preferred. (E,\$1) denotes the two-outcome act assigning \$1 to all states in E and \$0 to all remaining states.

Example 3.1 [A Bayesian Approach to Total Absence of Information]. The Bayesian approach requires assignment of probabilities P(E) to all uncertain events E. Some Bayesians analyze total absence of information by the "principle of insufficient reason," and set $P(s_1) = \ldots = P(s_n) = 1/n$. For decisions, such an analysis implies that $(s_1,\$1)$ is equivalent to $(s_2,\$1)$. Both acts are strictly preferred less than $(\{s_3,s_4\},\$1)$, $\{s_3,s_4\}$ being more likely than s_1 and than s_2. □

In the preceding example, the preferences depend on the particular way in which the state space was defined and invokes the counting measure on the state space. Events are more likely as they contain more states. The decisions would be different if state s_1 were split up into two states, say "s_1-and-heads" and "s_1-and-tails." Many have argued that this Bayesian approach does not reflect total absence of information, information being extracted from the counting measure on S.[14,15] More drastic forms of handling total absence of information have been proposed, a variant of which I will define next under the name principle of complete ignorance (PCI). First the principle is described, informally, in terms of "truth values." As this text adheres to the principles of decision theory and assumes decision making as the empirical primitive, the PCI will subsequently be formalized in terms of a decision principle.

A statement "$\{s_1,s_2\}$ is bigger/more likely than $\{s_1\}$" is no more accepted as meaningful. The *PCI, focused on S*, distinguishes only the following three states of information, or "truth values," for an event E.

- If E = S then E is *certain*.
- If E = ∅ then E is *impossible*.

In all other cases, E is uncertain, i.e.:

- If ∅ ≠ E ≠ F then E is *uncertain*.

The PCI is thus based on a three-valued logic. There are no different degrees of uncertainty and in this sense the processing of information is crude. If E is not impossible, then we also call E *possible*. For the PCI focused on S, we call S the *focal event* of complete ignorance. We also consider cases where the PCI is

---

[1] It can be considered the special case of the Dempster approach where there is only one possible random message at the first stage that is received with probability one.

focused on a subset F of S. This can occur for instance if the information has been received that F is true. In this case, parts of events outside of F are ignored. The *PCI, focused on* F, distinguishes the following states of information.

- If $E \supset F$, then E is *certain.*
- If $E \cap F = \varnothing$, then E is *impossible.*

In all other cases, E is uncertain, i.e.:

- If $\varnothing \neq E \cap F \neq F$, then E is *uncertain.*

The state of uncertainty is completely characterized, under the PCI, by the focal event. Before giving a definition in terms of preference conditions, let us discuss, in the next 1.5 page, an example of decision making that is in line with the PCI and that brings to the fore the most critical issue of the PCI.

Example 3.2 (PCI and Revealed Preference). We assume at least three states of nature in this example, so $n \geq 3$. Imagine that a choice must be made between the acts $(s_1,\$1)$ and $(s_2,\$1)$. For both acts the truth value of the outcome is the same. Under the left act, the event of receiving \$1 is the event s_1, which is possible but not certain, i.e. this event is uncertain. Under the right act, the event of receiving \$1 is $\{s_2\}$ which is again uncertain. For both acts, any outcome set containing neither \$1 nor \$0 is impossible, any outcome set containing both \$1 and \$0 is certain, and any outcome set containing either \$1 or \$0 but not both is uncertain.

Therefore, the PCI assumes no preference for either act. I will interpret this as an equivalence between the two acts, in agreement with the principles of revealed preference. Most authors in the domain of imprecise probabilities favor a different interpretation, where the above preference is indeterminate rather than equivalence. They argue that the above choice is not based on sufficient information and therefore does not reveal the value system of the decision maker, and no inference on preference should be based on it. This viewpoint underlies Cohen and Jaffray's analysis[14] as well as many upper and lower probability models[16–18], Walley[19] (Section 5.6), Walley[15] (p. 53 in reply to Lindley's objection to indecision). The disadvantage of this approach is that preference is disconnected from its observable basis, i.e. choice. A new empirical primitive must be introduced, so as to distinguish between choices revealing preferences and choices not revealing preferences. The new empirical primitive may be verbal communication with the decision maker or introspection. These are useful tools in practical decision aiding, and there are many psychological investigations into these phenomena. In a different context, Kahneman, Wakker and Sarin[20] have argued for a greater role of such tools in economics. In the present paper, however, I will base the analysis solely on the concepts of decision theory and revealed preference. Therefore, I assume equivalence

$$(s_1,\$1) \sim (s_2,\$1)$$

for the PCI.

Let me emphasize that completeness of preference is not required in this paper. Incompleteness of preference is permitted in the sense that not for all pairs of acts {f,g}, a choice between them needs to be considered. Once a choice situation is considered, however, a choice is compulsary and indeterminate preference is not allowed for (Wakker[21], Section III.1).

In the derived equivalence, the PCI does not yet deviate from the Bayesian principle of insufficient reason. However, the PCI also implies the following equivalence:

$$(s_1,\$1) \sim (\{s_2,s_3\},\$1).$$

Again, the same truth values of the outcomes result from both acts, with receipt of \$1 uncertain, etc. In this equivalence, the PCI deviates from the Bayesian principle of insufficient reason. The PCI does not accept cardinal information about sets and neither distinctions in size between different uncertain events (in agreement with "noninfluence of formalization",[14] or the "principle of representation invariance",[15] and deviating from some objective Bayesian approaches). Hence no claim is made that the event $\{s_2, s_3\}$ be larger than the event $\{s_1\}$.

A problematic feature of the PCI, and in my opinion its most critical property, appears from the following equivalence.

$$(s_2,\$1) \sim (\{s_2,s_3\},\$1).$$

The equivalence follows from the PCI as in the above reasonings because for both acts all outcome events have the same truth value. (It could also be derived from the preceding two equivalences through transitivity.) However, the right act dominates the left act and most people will strictly prefer the right act to the left act. The information processing of the PCI is too crude to detect statewise monotonicity as appearing here.

The violation of strict statewise monotonicity may be less problematic than seems at first sight. Also under expected utility, acts can be equivalent even though one always yields at least as much as the other in all states and strictly more in some states. Then expected utility will say that the latter states constitute a "null event." Similarly, the PCI can be defended by arguing that, if \$1 is received on event s_1, then adding event s_2 to that event does not change the truth value and in that sense s_2 can be considered a null event.

In summary, the PCI implies a violation of some strict monotonicity conditions which is, I think, the most critical aspect of the PCI. The claim of this paper is that decisions can (or cannot) be based on Dempster belief functions to the degree that the PCI as explained here is considered (un)reasonable.

□

Examples of decision principles that agree with the PCI are maximax and maximin decision making.[22] Cohen and Jaffray consider a somewhat different approach to the PCI.[14] Whenever there is a conflict between the above PCI and strict monotonicity, priority is given to strict monotonicity. As the authors point out, their approach necessarily requires violations of transitivity. Their approach also does not agree with the use of belief functions as does the PCI of this paper.

We discussed "truth values" without yet formalizing it. A decision-theoretic formalization should be in terms of preferences over acts. As a preparation for such a formalization, we relate the uncertainty about S to uncertainty about the outcomes of acts. Assume that $F \subset S$ is the focal event of complete ignorance and d is an act under consideration. The following can be said about the event of outcomes being contained in B, for any subset B of $\mathcal{C}$. Remember that by traditional conventions of measure theory, given act d, the event of receiving an outcome in $B \subset \mathcal{C}$ is *identified* with the event $d^{-1}(B) \subset S$.

- If $B \supset d(F)$, then B is *certain*;
- If $B \cap d(F) = \varnothing$, then B is *impossible*;

In all other cases, B is uncertain, i.e.:

- If $\varnothing \neq B \cap d(F) \neq d(F)$, then B is *uncertain.*

That is, for d the state of information on the outcome set can be described as complete ignorance focused on d(F). Under complete ignorance, the three truth values give a complete, "sufficient," description of the degree of uncertainty that is relevant for the evaluation of an act. Nothing else regarding the uncertainty about the state space is considered relevant. Thus, if for two different acts d and d', the truth values generated by the two acts coincide on the entire outcome space, then the acts should be equivalent in every respect regarding their preference value. This occurred in Example 3.2 for the acts $(s_2,\$1)$ and $(\{s_2,s_3\},\$1)$. Under both acts, any outcome set that contains both 1 and 0 is certain, any outcome set that contains neither 1 nor 0 is impossible, and the remaining outcome sets, that either contain 1 or 0 but not both, have truth value uncertain. In other words, both acts generate complete ignorance focused on $\{0,1\}$ over the outcome set $\mathcal{C}$.

Principle of complete ignorance (PCI), *focused on* the event $F \subset S$: Acts d and d' are equivalent whenever $d(F) = d'(F)$. □

Complete ignorance is completely characterized by its focal event F. Given an act d, the information regarding the outcome can then be described as complete ignorance focused on d(F). Under some natural preference conditions, including a weak monotonicity condition, it can be proved that the only decision making compatible with the PCI is decision making where an act d is evaluated by $U(\max(d(F)),\min(d(F)))$. This result will not be used in what follows, hence no proof is given.

In summary, the PCI permits complete objectivity by using only a minimal amount of information and deviates from the Bayesian principles. It is too crude to detect cases of statewise monotonicity and therefore violates strict monotonicity.

4. Dempster Belief Functions Derived from the Principle of Complete Ignorance

In this section, a decision-theoretic basis for belief functions is given that brings to the fore the role of the PCI. We define a decision model based on the two-stage model of uncertainty described in the introduction and add some assumptions to the PCI. I think that the added assumptions are relatively uncontroversial and that the critical assumption underlying Dempster belief functions is the PCI. That is, Dempster belief functions are appropriate if and only if the PCI, with its violation of strict monotonicity, is accepted. We turn to the defense of that claim. Because the PCI is a trivial case of the belief-function approach (with only one possible random message), we will concentrate on the derivation of belief functions from the PCI.

We assume the same decision model as before but assume that a specific kind of information is available on the state space, through random messages. (Hence there is no total absence of information on S, contrary to the preceding section.) The resulting two-stage model is depicted in Figure 1. This two-stage modeling of uncertainty is especially useful in the modeling of incomplete data.[23] Other two-stage models with uncertainty probabilized in one stage but not the other are considered in statistics and in the decision model of Anscombe and Aumann[24]. Figure 1 depicts an act d assigning d(s) = c to state s.

The information about which state of nature is true is somewhat complex and is described by a message space Ω. A message will be received in stage 1 which specifies a subset of S that contains the true state of nature. But the decision maker does not know beforehand what message he will receive. He knows that the message is an element of the message space Ω but is uncertain which element of Ω it is. This uncertainty is probabilized, that is, a probability measure π on Ω describes the probability distribution regarding which message will be received.

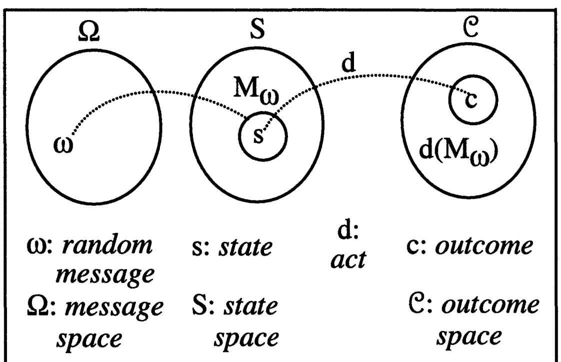

FIGURE 1 (The two-stage model)

For each possible message ω, a subevent M_ω of S is specified. The decision maker knows that, if ω is received, then the true state of nature is contained in M_ω. He does not have other information, so given M_ω he faces total absence of information. Because the decision maker is uncertain about which message he will receive, he is uncertain what the subset M_ω is. Obviously, if ω is the message received and the decision maker has chosen act d, then the resulting outcome will be an element of $d(M_\omega)$. In this model, the state space does not specify all uncertainty involved because it does not specify the random message received. One state of nature can be combined with different random messages. This is the characteristic property of the Dempster model. Let me emphasize that total absence of information does not concern S, but M_ω conditional on receipt of message ω.

We consider both *posterior preferences* $\succcurlyeq_\omega$ over acts, pertaining to choices between acts made after the receipt of a random message ω, and (*prior*) *preferences* $\succcurlyeq$ over acts, pertaining to choices made prior to that receipt. Throughout, unqualified preferences are understood to be prior, and so are statements about equivalence. I will argue that (prior) preferences are based on belief functions if posterior preferences satisfy the PCI.

Throughout, the outcome that will result for the decision maker is completely determined by the act d chosen by the decision maker and the true state of nature s. The only impact of the message ω on the outcome is "through" the true state s. Given the true state of nature s, the message ω does not have any more impact on the outcome of an act d.

We can now define a belief function on S. The decision principles introduced later will imply that this belief function comprises all the information regarding uncertainty that is relevant for (prior) decision making. For now, the belief function is a mathematical construct without yet any claim about empirical or decision-theoretic content. First note that the probability measure π and the mapping $\omega \mapsto M_\omega$ generate probabilities on 2^S, the collection of all subsets of S, in the natural manner. That is, $\varphi(E) = \pi(\omega: M_\omega = E)$ describes the probability that the message received will specify E as the event containing the true state of nature. We can now define the belief function f through φ as described in Section 1, i.e. $f(E) = \sum_{E' \subset E} \varphi(E')$. f(E) is precisely the probability that the random message implies a subset of E, hence certainty of E. The belief function describes, in short, the "probability of certainty." Other similar terms are probability of provability,[25] probability of knowing,[26] and probability of necessity[27].

Next we describe two decision principles (where the second strengthens the first) that imply that decisions must be based on the belief function. The first principle adapts the PCI for posterior preferences to prior preferences. The PCI is now applied conditionally given each ω.

Principle 1 (*prior PCI*). If, for each $\omega \in \Omega$, $d(M_\omega) = d'(M_\omega)$, then d and d' are equivalent. □

This principle reduces to the PCI if there is only one ω. Consider the posterior situation where the decision maker has received the message ω. Then the equality $d(M_\omega) = d'(M_\omega)$ implies, by the PCI, that d and d' are equivalent. Moreover, then d and d' generate exactly the same information regarding the outcome, i.e. each outcome set $B \subset \mathcal{C}$ has the same truth value (certain, impossible, or uncertain) under d as under d'. If the information regarding the outcomes generated by d and d' is the same for each ω, then prior equivalence between d and d' is required. To emphasize the elementary nature of Principle 1, let me display, and discuss in some detail, another condition that is quite stronger and is not needed. In the discussion of this principle we assume, for simplicity, that $\succcurlyeq$ and each $\succcurlyeq_\omega$ are weak orders (transitive and complete). Hence, equivalence coincides with the symmetric parts of $\succcurlyeq$ and $\succcurlyeq_\omega$, denoted by $\sim$ and $\sim_\omega$.

(Principle, not valid for belief functions) If, conditional on each $\omega \in \Omega$, d $\sim_\omega$ d' (posterior equivalence), then d $\sim$ d' (prior equivalence). □

The agreement of prior and posterior preference just displayed resembles somewhat the "dynamic consistency" condition from dynamic decision under risk.[28] In rich models, where each event can occur in a first and also second stage, the latter condition comprises a nontrivial part of the "separability" or "independence" preference condition that characterizes Bayesianism. Such a logic is not assumed in our defense of the prior PCI.

Our defense is as follows. The prior PCI assumes that the uncertainty about the generated outcome is identical for d and d', given each ω. If the uncertainty-information is identical for each ω, then it is also identical prior to the receipt of ω. Finally, only as a consequence of identical uncertainty regarding the resulting outcome, d and d' are required to be equivalent.

The prior PCI does not impose consistency between prior and posterior preference, but between prior and posterior identity of information. It can be considered a principle of logic rather than of belief. It is similar to an elementary and usually implicit assumption of decision under risk: that two options are equivalent or even identical if they generate the same tree, i.e. the same branches with the same probabilities at branches or outcomes at each stage. Our condition is of the same kind. Two options are equivalent if they generate the same branches with the same complete ignorance focused on the same branches or outcomes at each stage.

As a preparation for the second principle, we reformulate the first principle: d and d' are equivalent whenever, for each $B \subset \mathcal{C}$,

$$\{\omega \in \Omega: d(M_\omega) = B\} \;=\; \{\omega \in \Omega: d'(M_\omega) = B\}.$$

We now turn to the second principle. It reinforces the first by assuming that the only relevant aspect of the ωs is the probability mass they carry and that other than that their identity is not relevant. This is typically the assumption underlying decision under risk, where only probabilities are relevant and not the events generating them. We impose this assumption on the probabilized first-stage uncertainty.

Principle 2 (*neutrality axiom*). Acts d and d' are equivalent whenever, for each $B \subset \mathcal{C}$,

$$\pi\{\omega \in \Omega: d(M_\omega) = B\} = \pi\{\omega \in \Omega: d'(M_\omega) = B\}. \quad \square$$

This principle characterizes the relevance of belief functions for decision making. That is, the preference value of an act is completely determined by the belief function it generates over the outcome set. In the following theorem, f is the belief function on S defined before and $f \circ d^{-1}$ is the belief function on the outcome set assigning $f(d^{-}(B))$ to each $B \subset \mathcal{C}$.

Theorem 4.1. Neutrality holds if and only if, for all acts d, d':
$d \sim d'$ whenever $f \circ d^{-1} = f \circ d'^{-1}$.

Proof. Consider the following four equalities, each imposed on all $B \subset \mathcal{C}$, and discussed next.

$$\pi\{\omega: d(M_\omega) = B\} = \pi\{\omega: d'(M_\omega) = B\};$$

$$\pi\{\omega: d(M_\omega) \subset B\} = \pi\{\omega: d'(M_\omega) \subset B\};$$

$$\pi\{\omega: M_\omega \subset d^{-1}(B)\} = \pi\{\omega: M_\omega \subset d'^{-1}(B)\};$$

$$f(d^{-1}(B)) = f(d'^{-1}(B)).$$

Equivalence of the first two equalities can be proved by induction with respect to the number of elements of B, equivalence of the second and third equalities follows from elementary set-theory, and equivalence of the last two equalities follows from the definition of the belief function f. Neutrality requires that the first equality, for all B, imply that d and d' are equivalent, the second part of Theorem 4.1 requires the same implication for the fourth equality. By the equivalence of the first and fourth equalities, the theorem follows. $\square$

Under neutrality, all the information about the uncertainty regarding S and Ω relevant for decision making is apparently captured by the belief function f. In statistical terminology, the belief function provides a "sufficient" description of the uncertainty. The neutrality axiom has thus provided a decision-theoretic foundation for belief functions in decision making, based on the PCI. Other than that, the theorem leaves complete freedom regarding the manner in which decisions are derived from belief functions. It does not impose transitivity or completeness.

It is logically possible that prior preferences are based on a belief function, so are *as if* based on the PCI and its extensions, but that preferences after actual receipt of a random message are different and do not comply with the PCI. Dynamic consistency principles could be formulated to rule such cases out. We do not pursue this issue further.

The result of this paper holds within the structure depicted in Figure 1. The belief function on S cannot be arbitrary but must be the one generated by the depicted structure. The PCI must also be as depicted, focused on M_ω for each ω. An extreme case arises if each M_ω is a singleton. Then our approach yields only trivial degenerate cases of the PCI for each ω, and the belief function resulting on S is simply the additive Bayesian measure generated by π. Deviations from Bayesianism occur when at least one M_ω is not a singleton.

Let me emphasize that obtaining a preference axiomatization for belief functions is not the purpose of this paper. The belief function is objectively given, hence a preference axiomatization is trivial. Preference axiomatizations are only nontrivial if some quantities are not given a priori, such as in Ghirardato[6]. The purpose of this paper is to show the relations between belief functions and the PCI.

5. Summary and Conclusion

Imagine that decisions must be made while facing uncertainty, and the uncertainty is resolved in two stages. The first-stage uncertainty can be probabilized but the second cannot. Imagine that the decision maker does not want to deal with the second-stage nonprobabilized uncertainty in a Bayesian manner, but instead wants to follow the principle of complete ignorance, e.g. so as to preserve complete objectivity of the decision procedure. Then, as is the claim of this note, the decisions necessarily go by belief functions. So as to establish this claim, the principle of complete ignorance was reinforced, first, to the prior principle of complete ignorance, second, to the neutrality principle. These reinforcements seem relatively uncontroversial and can be compared to conditions implicitly assumed in decision under risk. Hence the crucial step from Bayesianism to Dempster belief functions seems to be the adoption of the principle of complete ignorance.

Acknowledgment

Many of the ideas presented in this paper have been obtained through communications with Jean-Yves Jaffray. The paper does, however, express my personal opinions that need not agree with Jaffray's, and the usual disclaimer applies. Two anonymous referees, Philippe Smets, and Mark Machina gave helpful comments.

References

1. P. Smets, "Constructing the pignistic probability function in a context of uncertainty", in *Uncertainty in artificial Intelligence* 5, eds. M. Henrion, R.D. Shachter, L.N. Kanal, and J.F. Lemmer (North-Holland, Amsterdam, 1989) pp. 29–40.
2. J.Y. Jaffray, "Linear utility theory for belief functions", *Oper. Res. Lett.* **8** (1989a) 107–112.

3. J.Y. Jaffray, "Dynamic decision making with belief functions", in *Advances in the Dempster-Shafer Theory of Evidence*, eds. R.R.Yager, M. Fedrizzi, and J. Kacprzyk (Wiley, New York, 1994) pp. 331–352.
4. T.M. Strat, "Decision analysis using belief functions", *Internat. J. Approx. Reas.* **4** (1990) 391–418.
5. P. Smets and R. Kennes, "The transferable belief model", *Artificial Intelligence* **66** (1994) 191–234.
6. P. Ghirardato, "Coping with ignorance: unforeseen contingencies and non-additive uncertainty", Working Paper, Caltech Social Science, Pasadena CA, 1999.
7. I. Gilboa and D. Schmeidler, "Canonical representation of set functions", *Math. Oper. Res.* **20** (1995) 197–212.
8. E. Hendon, H.J. Jacobsen, B. Sloth, and T. Tranaes, "Expected utility with lower probabilities", *J. Risk Uncert.* **8** (1994) 197–216.
9. S. Mukerji, "Understanding the nonadditive probability decision model", *Econom. Theory.* **9** (1996) 23–46.
10. A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Ann. Math. Statist.* **38** (1967) 325–339.
11. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton NJ, 1976).
12. J.Y. Jaffray and P.P. Wakker, "Decision making with belief functions: compatibility and incompatibility with the sure-thing principle", *J. Risk Uncert.* **8** (1993) 255–271.
13. L.J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954). (Second edition Dover, New York, 1972.)
14. M. Cohen and J.Y. Jaffray, "Rational behavior under complete ignorance", *Econometrica* **48** (1980) 1281–1299.
15. P. Walley, "Inferences from multinomial data: learning about a bag of marbles", *J. Roy. Statist. Soc. Ser. B* **58** (1996) 3–57.
16. C.A.B. Smith, "Consistency in statistical inference and decision", *J. Roy. Statist. Soc. Ser. B* **23** (1961) 1–25.
17. I.J. Good, *Good Thinking*. (University of Minnesota Press, Minneapolis, MN, 1983).
18. H.E. Kyburg, Jr., *Epistemology and Inference* (University of Minnesota Press, Minneapolis, 1983).
19. P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991).
20. D. Kahneman, P.P. Wakker, & R.K. Sarin, "Back to Bentham? Explorations of experienced utility", *Quart. J. Econom.* **112** (1997) 375–405.
21. P.P. Wakker, *Additive Representations of Preferences, A New Foundation of Decision Analysis* (Kluwer Academic Publishers, Dordrecht, 1989).
22. K.J. Arrow and L. Hurwicz, "An optimality criterion for decision making under ignorance", in *Uncertainty and Expectations in Economics*, eds. C.F. Carter and J.L. Ford (Basil Blackwell and Mott Ltd., Oxford, UK, 1972).
23. J.Y. Jaffray, "Coherent bets under partially resolving uncertainty and belief functions", *Theory and Decision* **26** (1989b) 99–105.
24. F.J. Anscombe and R.J. Aumann, "A definition of subjective probability", *Ann. Math. Statist* **34** (1963) 199–205.
25. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo CA, 1988).

26. E.H. Ruspini, "The logical foundations of evidential reasoning, revised version", Technical note 408, SRI International, Menlo Park, CA, USA, 1987.
27. D. Dubois and H. Prade, "Default reasoning and possibility theory", *Artificial Intelligence* **35** (1988) 243–257.
28. M.J. Machina, "Dynamic consistency and non-expected utility models of choice under uncertainty", *J. Econom. Lit.* **27** (1989) 1622–1688.