INDUCTIVE INFERENCE AND NONMONOTONIC REASONING

Marco E. G. V. Cattaneo

March 2000

Diplomarbeit (Master's Thesis)

written under the guidance of

Prof. Frank Hampel

Seminar for Statistics
Department of Mathematics

Swiss Federal Institute of Technology Zürich

Contents

1 An extension of classical propositional logic which allows inductive inferences

1.1 Introduction

Symbolic logic is the discipline studying mathematical models of correct human reasoning. I do not want to discuss here the meaning of "correct", because this will lead to endless philosophical debates, but I would like to formulate two important remarks about the first sentence. First, since mathematics is a form of (correct) human reasoning, we have to distinguish between the reasoning studied and the one employed in this study: we can do this by expressing the first in a rigorously defined symbolic language (called the object-language) and the latter in our usual (mathematical) language (called the metalanguage). Second, the aim of a mathematical model is to represent a simplified reality: you cannot model something without losing generality and precision, but this loss can be partially balanced by simplicity and mathematical beauty.

Propositional (symbolic) logic studies the connections between propositions: a proposition is a declarative (i.e., not interrogative or imperative) sentence, which we agree to view as being either true or false. I have written "which we agree to view as" because all sentences of a natural language have ambiguity: we could try to reduce this ambiguity, but we shall always have to rely on our common understanding. For instance I think that "Socrates is a man" is an acceptable proposition, even if someone could ask for a precise definition of "being a man", whereas an ambiguous statement such as "Socrates is old" is not a proposition, until "old" is precisely defined. An example of a sentence which surely is not a proposition is "This sentence is false" (since it can be neither true nor false), whereas the sentence "The event A has the probability p" will be a proposition only if we agree that probability is an univocal characteristic of an event like A. In propositional logic the internal structure of propositions is ignored except insofar as they are built from other considered propositions using specified connectives. For instance "If Socrates is a man, then Socrates is mortal" is constructed from the simpler propositions "Socrates is a man" and "Socrates is mortal" using the connective "If..., then...". Of course if we ignore the internal structure of a proposition, we have to ignore its meaning too: what remains of the semantics in propositional logic are merely the logical connectives. This is a huge limitation, but allows the great result of classical propositional logic: to find every correct conclusion from a set of premises by analyzing only the syntactical level. For instance from "If Socrates is a man, then Socrates is mortal" and "Socrates is a man" it is correct to conclude "Socrates is mortal", without considering the meaning of these propositions (but only the meaning of the connective "If..., then..."), whereas from "Socrates is a man" it is incorrect to conclude "Socrates is mortal" without considering the semantics of these sentences.

Classical propositional logic studies the sure consequences of certainties; i.e., what we are compelled to accept as a certainty if we are sure of the premises and we reason correctly. That is, classical propositional logic models deductive inference, where "deductive inference" is not to be interpreted with the classical meaning of "reasoning leading from general premises to particular conclusions", but with the present-day meaning of "inference whose validity is indubitable". In classical terminology, the opposite of deductive is inductive (i.e., "leading from particular premises to general conclusions"), whereas the present-day meaning of "inductive inference" is not precise: inductive is an inference which leads to conclusions that are very likely but in general not sure. Classical propositional logic cannot express inductive inferences, but these are fundamental in human reasoning. To allow inductive inferences in propositional logic it is necessary to generalize the classical formulation; there are several ways of doing this, but all have a common feature: nonmonotonicity. Deductive reasoning is monotonic in the sense that, having derived a conclusion from some believed premises, nothing we can add to our knowledge will change that conclusion, given that we still hold the premises (but we shall see that in fact this is true only because of an unjustified particularity of the mathematical model). Because of its fallibility, inductive reasoning does not share this property: new knowledge can compel us to withdraw some of our former conclusions.

There are fundamentally two ways to extend the classical formulation so as to allow inductive inferences in propositional logic: by merely extending the notion of logical consequence or by generalizing the notion of knowledge. The first is the way followed by default logic, but the results are unsatisfactory: for every specific situation the possible inductive inferences have to be declared. The second is the way I shall follow: generalizing the notion of knowledge by allowing beliefs that are not certainties, without changing anything at the ontological level: to remain in the field of propositional logic we have to accept that propositions can be either true or false, but this does not prevent us to believe that a proposition is probable or improbable. For instance I can believe that the sentence "In Lucerne it is raining" is very likely, without calling in question that in fact it is true or false. The notion of logical consequence will also be extended (to allow us to accept a very likely proposition), but this in a very natural and general way: in particular, default logic can be considered as a special case of the formulation presented. As in the classical formulation (in which we can only be certain of a proposition or say nothing about it), we can only support a proposition: not supporting it does not mean impugning it, to impugn it we have to support its negation. Therefore the allowed beliefs are support-values: we can extend the notion of knowledge by allowing further support-values besides the classical "certain". We could do this in a discrete way (for instance allowing the support-values "certain", "almost certain", "very likely", "likely", "quite likely",...), but I find more natural (and mathematically more elegant) to do it in a continuous way: to allow as support-values the real numbers in the unit interval $[0, 1]$ (where 1 means "certain" and 0 is the equivalent of saying nothing in the classical formulation), even if it can be difficult to understand the meaning of a number in $(0, 1)$ (but the discrete case can be seen as a particular case, by assigning to each one of the considered support-values a real number). Someone could think that this extension of the notion of belief to the unit interval leads to (epistemic) probabilities, but it is important to notice that probability theory is not an extension of classical propositional logic (because in probability theory not supporting a proposition means impugning it); we shall see that in fact Bayesian probability is a particular case of the theory presented.

I shall now present the static aspect of the theory (the representation of a belief and the conclusions drawn) and later the dynamic one (the evolution of a belief): in both cases this theory stays between the two incompatible special cases of classical propositional logic and Bayesian probability.

1.2 Syntax

Propositional logic deals with formal languages (object-languages) whose elements, called formulas, are particular finite strings on an alphabet consisting of propositional symbols (which represent arbitrary propositions), constants (like "truth" and "falsehood"), logical connectives (like "and", "or", "not",...) and punctuation marks (like brackets). I shall denote sets of propositional symbols with $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$,... and their elements with small Latin letters such as a, b, c,.... I consider only finite sets (possibly empty) of propositional symbols, because I do not think that infinite sets of propositions have something to do with human reasoning: a set of propositional symbols represents a particular topic. As constants I consider $\top$ (truth) and $\bot$ (falsehood), as logical connectives $\neg$ (negation), $\vee$ (disjunction), $\wedge$ (conjunction), $\rightarrow$ (implication) and $\leftrightarrow$ (equivalence), and as punctuation marks the brackets (and). All the others mathematical symbols that I shall use are to be interpreted as elements of the metalanguage, and the only symbols in common between the object-language and the metalanguage will be the brackets and the small Latin letters. The symbol $\equiv$ means "is the string", for instance $\varphi \equiv (a \rightarrow b)$ signifies that φ is a symbol in the metalanguage for the string $(a \rightarrow b)$ of the object-language.

Definition 1 *Given a set of propositional symbols $\mathcal{U}$, the language $\mathcal{L}_{\mathcal{U}}$ (whose elements are called formulas on $\mathcal{U}$) is the least set $\mathcal{L}$ of strings over the alphabet*

$$\mathcal{U} \cup \{\top, \bot, \neg, \vee, \wedge, \rightarrow, \leftrightarrow, (,)\},$$

such that:

- $(\mathcal{U} \cup \{\top, \bot\}) \subset \mathcal{L}$;[1]
- $\varphi \in \mathcal{L} \Rightarrow \neg\varphi \in \mathcal{L}$;

[1] I interpret the symbol $\subset$ as "is a subset of" (other authors would use the symbol $\subseteq$); to indicate "is a proper subset of", I use the symbol $\subsetneq$.

- $\{\varphi, \psi\} \subset \mathcal{L} \Rightarrow \{(\varphi \vee \psi), (\varphi \wedge \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi),\} \subset \mathcal{L}$.

It is easy to show that this set always exists (and is countably infinite) and that $\mathcal{U} \subset \mathcal{V} \Rightarrow \mathcal{L}_{\mathcal{U}} \subset \mathcal{L}_{\mathcal{V}}$. Of course two (syntactically) different formulas can be semantically equivalent (for instance $\neg(a \wedge b)$ and $(\neg a \vee \neg b)$), in fact the alphabet used is redundant (but very handy): for example we could consider as sole constant $\top$ and as connectives only $\neg$ and $\vee$, and define the others symbols as abbreviations (in the metalanguage): for instance $\perp$ as abbreviation of $\neg\top$ (i.e., $\perp \equiv \neg\top$) and $(\varphi \rightarrow \psi)$ of $(\neg\varphi \vee \psi)$. Even in our redundant language abbreviations can be very useful, as it is the case for the following two.

Definition 2 *For a finite set $F = \{\varphi_1, \dots, \varphi_n\}$ of formulas, the formulas $\bigvee F$ and $\bigwedge F$ are recursively defined as follows:*

- $\bigvee \emptyset \equiv \perp$ *and* $\bigwedge \emptyset \equiv \top$*;*
- $\bigvee \{\varphi_1, \dots, \varphi_{i+1}\} \equiv (\bigvee \{\varphi_1, \dots, \varphi_i\} \vee \varphi_{i+1})$ *and*
  $\bigwedge \{\varphi_1, \dots, \varphi_{i+1}\} \equiv (\bigwedge \{\varphi_1, \dots, \varphi_i\} \wedge \varphi_{i+1})$.

Of course, given a finite set F of formulas, $\bigvee F$ (or $\bigwedge F$) can represent $|F|!$ different formulas[2], depending on the choice of a numbering of the elements of F, but all these formulas are semantically equivalent. For instance $\bigwedge \{a, (a \rightarrow b)\}$ can represent $((\top \wedge a) \wedge (a \rightarrow b))$ or $((\top \wedge (a \rightarrow b)) \wedge a)$, but of course both have the same meaning of $(a \wedge b)$.

In classical propositional logic we describe our belief on a topic $\mathcal{U}$ by declaring at most one formula[3] on $\mathcal{U}$: if a formula is declared, then the meaning is that we believe it to be "certain", otherwise we believe that the truth $\top$ is "certain" (i.e., we say nothing). In the new formulation the support-value "certain" is 1 and the only other number with a corresponding classical interpretation is 0, which means "no support". Therefore we can express the classical case by saying that we distribute the total support 1 by assigning 1 to at most one formula, 0 to some other formulas and the rest to the truth $\top$.

Definition 3 *Given a set of propositional symbols $\mathcal{U}$, a graded formula on $\mathcal{U}$ is an element of the set $\mathcal{G}_{\mathcal{U}} = \{\langle\varphi, \alpha\rangle \mid \varphi \in \mathcal{L}_{\mathcal{U}},\ \alpha \in [0, 1]\}$.*

[2]Given a finite set A, $|A|$ indicates its cardinality.

[3]In the standard notation of classical propositional logic it is also possible to declare more formulas, but this is completely equivalent to the declaration of their conjunction, which is a single formula. Therefore I will consider only the notation with at most one declared formula.

Definition 4 *Given a set of propositional symbols $\mathcal{U}$, a support distribution on $\mathcal{U}$ is a finite set $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\} \subset \mathcal{G}_{\mathcal{U}}$ such that $\sum_{i=1}^{n} \alpha_i \leq 1$ and $i \neq j \Rightarrow \varphi_i \not\equiv \varphi_j$.*

We interpret a support distribution $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ as the belief assigning α_i to φ_i (for i going from 1 to n) and the rest $(1- \sum_{i=1}^{n} \alpha_i)$ to the truth $\top$. This is compatible with the classical case, in which we consider only support distributions $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ such that $\alpha_i \in \{0, 1\}$ for every $i \in \{1, \ldots, n\}$ (in the following I shall call such support distributions "classical"). Of course different support distributions can describe the same support assignment: for instance both $\{\langle\varphi, \alpha\rangle\}$ and $\left\{\left\langle(\varphi \wedge \varphi), \frac{\alpha}{2}\right\rangle, \left\langle\varphi, \frac{\alpha}{2}\right\rangle\right\}$ assign α to φ and $1-\alpha$ to $\top$ (since $(\varphi \wedge \varphi)$ has obviously the same meaning of φ).

I have defined a support distribution on $\mathcal{U}$ as a finite subset of $\mathcal{G}_{\mathcal{U}}$: to allow infinite subset would not change anything in the theory, but I do not think that such a generalization would describe human reasoning better. I have put the condition $i \neq j \Rightarrow \varphi_i \not\equiv \varphi_j$, only to avoid tedious remarks in the following proofs: of course this is a restriction only from a purely formal standpoint. It is important to notice that if $\mathcal{U} \subset \mathcal{V}$ are two sets of propositional symbols, then, since of course $\mathcal{G}_{\mathcal{U}} \subset \mathcal{G}_{\mathcal{V}}$, a support distribution on $\mathcal{U}$ is also one on $\mathcal{V}$.

Example 5 *Consider the topic of the weather situation of Lucerne: for instance I could believe that it is very likely that it is raining in Lucerne, and that in this case it is likely that the wind is blowing. We can model this simple topic by considering the two propositions "In Lucerne it is raining" and "In Lucerne the wind is blowing", represented by the propositional symbols r and w, respectively; thus we consider a set of propositional symbols $\mathcal{U} \supset \{r, w\}$. I believe that r is very likely, that is, I support it a lot, say with 0.9; but, since I believe that in the case of rain the wind is likely, a large part of this support-value has to be given to $(r \wedge w)$, say 0.7. Therefore we can express this belief with the support distribution $\{\langle(r \wedge w), 0.7\rangle, \langle r, 0.2\rangle\}$ on $\mathcal{U}$.*

1.3 Semantics

As I have said before, a proposition (that is, a formula of the object-language) is either true or false: a valuation in propositional logic is an assignment of one of the two possible truth-values (t or f) to every proposition, such that the well-known truth-tables for the logical connectives are satisfied, and such that the truth-values of the constants $\top$ and $\bot$ are t and f, respectively. Proposition 6 states that, since every formula is built from the propositional symbols and the constants using the logical connectives, a valuation of $\mathcal{L}_{\mathcal{U}}$ is completely defined by a valuation of $\mathcal{U}$, which is an assignment of a truth-value to every element of $\mathcal{U}$ (the set of the valuations of $\mathcal{U}$, denoted $V_{\mathcal{U}}$, is $\{t,f\}^{\mathcal{U}}$)[4]. Therefore $\mathcal{L}_{\mathcal{U}}$ has $|V_{\mathcal{U}}| = 2^{|\mathcal{U}|}$ different valuations (this is correct even if $\mathcal{U} = \emptyset$: in fact, from naive set theory follows that $\{t,f\}^{\emptyset} = \{\emptyset\}$).

Proposition 6 *Given a set of propositional symbols $\mathcal{U}$ and a valuation v of $\mathcal{U}$ ($v \in V_{\mathcal{U}}$), there is a unique extension $\bar{v}$ of v to $\mathcal{L}_{\mathcal{U}}$ ($\bar{v} \in \{t,f\}^{\mathcal{L}_{\mathcal{U}}}$ and $\bar{v}{\downarrow_{\mathcal{U}}} = v$)[5] which is a valuation of $\mathcal{L}_{\mathcal{U}}$, that is:*

- $\bar{v}(\top) = t$ *and* $\bar{v}(\bot) = f$;
- $\varphi \in \mathcal{L}_{\mathcal{U}} \Rightarrow \bar{v}(\neg\varphi) \neq \bar{v}(\varphi)$;
- $\{\varphi, \psi\} \subset \mathcal{L}_{\mathcal{U}} \Rightarrow \begin{cases} \bar{v}((\varphi \vee \psi)) = t \Leftrightarrow (\bar{v}(\varphi) = t \text{ or } \bar{v}(\psi) = t), \\ \bar{v}((\varphi \wedge \psi)) = t \Leftrightarrow (\bar{v}(\varphi) = t \text{ and } \bar{v}(\psi) = t), \\ \bar{v}((\varphi \rightarrow \psi)) = f \Leftrightarrow (\bar{v}(\varphi) = t \text{ and } \bar{v}(\psi) = f), \\ \bar{v}((\varphi \leftrightarrow \psi)) = t \Leftrightarrow \bar{v}(\varphi) = \bar{v}(\psi). \end{cases}$

Proof. Trivially by induction on the number of logical connectives in a formula. ■

Given a formula φ on $\mathcal{U}$, a valuation $v \in V_{\mathcal{U}}$ is a model of φ if $\bar{v}(\varphi) = t$. The models of a proposition are thus the valuations that regard it as true, for instance $\bot$ has no models and every valuation is a model of $\top$. The mapping[6]

$$\begin{array}{rccl} s_{\mathcal{U}}: & \mathcal{L}_{\mathcal{U}} & \longrightarrow & \mathcal{P}(V_{\mathcal{U}}) \\ & \varphi & \longmapsto & \{v \in V_{\mathcal{U}} \mid \bar{v}(\varphi) = t\}, \end{array}$$

which assigns to every formula on $\mathcal{U}$ the set of its models, is called the semantical mapping of $\mathcal{U}$ (in propositional logic the semantics of a proposition is completely defined by the set of its models). Proposition 7 states that for every set of valuations of $\mathcal{U}$ there is a formula on $\mathcal{U}$ with exactly these valuations as models: the language $\mathcal{L}_{\mathcal{U}}$ is large enough to express every meaning accepted in propositional logic.

Proposition 7 *Given a set s of valuations of $\mathcal{U}$, there is a formula φ on $\mathcal{U}$ such that $s_{\mathcal{U}}(\varphi) = s$.*

Proof. Define $\varphi \equiv \bigvee \{\bigwedge (\{a \mid a \in \mathcal{U},\ w(a) = t\} \cup \{\neg a \mid a \in \mathcal{U},\ w(a) = f\}) \mid w \in s\}$. A valuation $v \in V_{\mathcal{U}}$ is a model of φ if $\bar{v}(\varphi) = t$;

[4] Given two sets A and B, A^B is the set of the mappings from B to A; i.e., $A^B = \{f \mid f : B \longrightarrow A\}$.

[5] Given three sets A, B and C, such that $B \subset C$, and a mapping $f \in A^C$, $f{\downarrow_B}$ is the restriction of f to B; i.e., $f{\downarrow_B} \in A^B$ and $\forall x \in B\ \ f{\downarrow_B}(x) = f(x)$.

[6] Given a set A, $\mathcal{P}(A)$ is the power set of A; i.e., $\mathcal{P}(A) = \{B \mid B \subset A\}$.

i.e., if $\bar{v}(\bot) = t$ (which is impossible) or if there is a $w \in s$ such that
$\bar{v}\left(\bigwedge\left(\{a \mid a \in \mathcal{U},\ w(a) = t\} \cup \{\neg a \mid a \in \mathcal{U},\ w(a) = f\}\right)\right) = t$;
the last expression is valid if $\bar{v}(\top) = t$ (which is always valid), $v(a) = t$ for every $a \in \mathcal{U}$ such that $w(a) = t$, and $v(\neg a) = t$ for every $a \in \mathcal{U}$ such that $w(a) = f$,
but this completely defines v: $v = w$.
Therefore $v \in V_{\mathcal{U}}$ is a model of φ if there is a $w \in s$ such that $v = w$; thus $s_{\mathcal{U}}(\varphi) = s$. ■

Proposition 8 *Given a set of propositional symbols $\mathcal{U}$, the semantical mapping of $\mathcal{U}$ is the unique mapping $s_{\mathcal{U}} \in \mathcal{P}(V_{\mathcal{U}})^{\mathcal{L}_{\mathcal{U}}}$ such that:*

- $a \in \mathcal{U} \Rightarrow s_{\mathcal{U}}(a) = \{v \in V_{\mathcal{U}} \mid v(a) = t\}$;
- $s_{\mathcal{U}}(\top) = V_{\mathcal{U}}$ *and* $s_{\mathcal{U}}(\bot) = \emptyset$;
- $\varphi \in \mathcal{L}_{\mathcal{U}} \Rightarrow s_{\mathcal{U}}(\neg\varphi) = V_{\mathcal{U}} \setminus s_{\mathcal{U}}(\varphi)$;
- $\{\varphi, \psi\} \subset \mathcal{L}_{\mathcal{U}} \Rightarrow \begin{cases} s_{\mathcal{U}}((\varphi \vee \psi)) = s_{\mathcal{U}}(\varphi) \cup s_{\mathcal{U}}(\psi), \\ s_{\mathcal{U}}((\varphi \wedge \psi)) = s_{\mathcal{U}}(\varphi) \cap s_{\mathcal{U}}(\psi), \\ s_{\mathcal{U}}((\varphi \to \psi)) = V_{\mathcal{U}} \setminus (s_{\mathcal{U}}(\varphi) \setminus s_{\mathcal{U}}(\psi)), \\ s_{\mathcal{U}}((\varphi \leftrightarrow \psi)) = V_{\mathcal{U}} \setminus ((s_{\mathcal{U}}(\varphi) \setminus s_{\mathcal{U}}(\psi)) \cup (s_{\mathcal{U}}(\psi) \setminus s_{\mathcal{U}}(\varphi))). \end{cases}$

Proof. Trivially by induction on the number of logical connectives in a formula. ■

As I have said before, in classical propositional logic we describe our belief on a topic $\mathcal{U}$ by declaring at most one formula on $\mathcal{U}$. If $\varphi \in \mathcal{L}_{\mathcal{U}}$ is declared, then we say that $\psi \in \mathcal{L}_{\mathcal{U}}$ is a semantical consequence of φ (in symbols $\varphi \models \psi$) if every model of φ is a model of ψ (i.e., if $s_{\mathcal{U}}(\varphi) \subset s_{\mathcal{U}}(\psi)$). If no formula is declared (which is equivalent to declaring $\top$), then we say that $\psi \in \mathcal{L}_{\mathcal{U}}$ is a tautology (in symbols $\models \psi$) if every valuation on $\mathcal{U}$ is a model of ψ(i.e., if $s_{\mathcal{U}}(\psi) = V_{\mathcal{U}}$; since $s_{\mathcal{U}}(\top) = V_{\mathcal{U}}$, a tautology is a semantical consequence of $\top$). If we describe a (classical) belief with a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\} \subset \mathcal{G}_{\mathcal{U}}$, a formula ψ on $\mathcal{U}$ is a semantical consequence of D (in symbols $D \models \psi$) if[7]

$$\sum_{i=1}^{n} \alpha_i \delta^0_{|s_{\mathcal{U}}(\varphi_i) \setminus s_{\mathcal{U}}(\psi)|} + \left(1 - \sum_{i=1}^{n} \alpha_i\right) \delta^0_{|V_{\mathcal{U}} \setminus s_{\mathcal{U}}(\psi)|} = 1;$$

that is, if the sum of the support-values assigned to the formulas of which ψ is a semantical consequence reaches its maximum 1. With this description it is easy to generalize the

[7] If m and n are two integers, δ^m_n is the Kronecker-delta; i.e., $\delta^m_n = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases}$ In particular, if A and B are two sets, then $\delta^0_{|A \setminus B|} = \begin{cases} 1 & \text{if } A \subset B, \\ 0 & \text{if } A \not\subset B. \end{cases}$

classical notion so as to allow uncertain inferences, by saying that ψ is a semantical consequence with grade $\alpha \in [0,1]$ of D (in symbols $D \overset{\alpha}{\models} \psi$) if

$$\sum_{i=1}^{n} \alpha_i \delta^0_{|s_{\mathcal{U}}(\varphi_i) \setminus s_{\mathcal{U}}(\psi)|} + \left(1 - \sum_{i=1}^{n} \alpha_i\right) \delta^0_{|V_{\mathcal{U}} \setminus s_{\mathcal{U}}(\psi)|} \geq \alpha;$$

that is, if the sum of the support-values assigned to the formulas of which ψ is a (classical) semantical consequence reaches the grade α[8]. So classical propositional logic becomes the particular case in which we allow only classical support distributions and we consider only semantical consequences with grade 1 (i.e., sure conclusions). Therefore if D is a classical support distribution, then $D \models \psi \Leftrightarrow D \overset{1}{\models} \psi$.

Consider two sets of propositional symbols $\mathcal{U}$ and $\mathcal{V}$ (which do not have to be disjoint): a support distribution on $\mathcal{U}$ is also one on $\mathcal{U} \cup \mathcal{V}$; so we can generalize the notion of semantical consequence by allowing formulas of other languages: a formula ψ on $\mathcal{V}$ is a semantical consequence with grade α of a support distribution D on $\mathcal{U}$ (in symbols $D \overset{\alpha}{\models} \psi$) if

$$\sum_{i=1}^{n} \alpha_i \delta^0_{|s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) \setminus s_{\mathcal{U}\cup\mathcal{V}}(\psi)|} + \left(1 - \sum_{i=1}^{n} \alpha_i\right) \delta^0_{|V_{\mathcal{U}\cup\mathcal{V}} \setminus s_{\mathcal{U}\cup\mathcal{V}}(\psi)|} \geq \alpha.$$

This definition is not very handy: we can simplify it by observing that actually we are not interested in the syntactical form of the supported propositions φ_i, but only in the sets of their models $s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)$; thus the following definition can be very useful.

Definition 9 *Given a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$, the associated semantical support distribution (SSD) on $\mathcal{U}$ is the function*

$$\begin{array}{rcl} m^D_{\mathcal{U}} : \ \mathcal{P}(V_{\mathcal{U}}) & \longrightarrow & [0,1] \\ s & \longmapsto & \begin{cases} \sum\limits_{s_{\mathcal{U}}(\varphi_i) = s} \alpha_i & \text{if } s \neq V_{\mathcal{U}}, \\ 1 - \sum\limits_{s_{\mathcal{U}}(\varphi_i) \neq V_{\mathcal{U}}} \alpha_i & \text{if } s = V_{\mathcal{U}}. \end{cases} \end{array}$$

That is, the SSD assigns to every set of valuations the sum of the supports given to formulas which have exactly these valuations as models (obviously $\sum\limits_{s \subset V_{\mathcal{U}}} m^D_{\mathcal{U}}(s) = 1$). Now it is possible to simplify the definition of semantical consequence.

Definition 10 *A formula ψ on $\mathcal{V}$ is a semantical consequence with grade α of a support distribution D on $\mathcal{U}$ (in symbols $D \overset{\alpha}{\models} \psi$) if $\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \alpha$.*

[8] From now on, with "grade" it will be implicit "$\in [0,1]$".

Example 11 *Consider the situation of Example 5: we had a support distribution* $D = \{\langle (r \wedge w), 0.7\rangle, \langle r, 0.2\rangle\}$ *on* $\mathcal{U} \supset \{r, w\}$. *For the sake of simplicity, let* $\mathcal{U} = \{r, w\}$; *thus* $V_{\mathcal{U}}$ *contains the four elements* v_{xy} *(for* $x, y \in \{t, f\}$*) such that* $v_{xy}(r) = x$ *and* $v_{xy}(w) = y$. *With this notation,* $s_{\mathcal{U}}((r \wedge w)) = \{v_{tt}\}$ *and* $s_{\mathcal{U}}(r) = \{v_{tt}, v_{tf}\}$*: therefore* $m_{\mathcal{U}}^{D}(\{v_{tt}\}) = 0.7$, $m_{\mathcal{U}}^{D}(\{v_{tt}, v_{tf}\}) = 0.2$ *and* $m_{\mathcal{U}}^{D}(V_{\mathcal{U}}) = 0.1$ *(*$m_{\mathcal{U}}^{D}(s) = 0$ *otherwise). Since* $\sum_{s \subset s_{\mathcal{U}}(r)} m_{\mathcal{U}}^{D}(s) = \sum_{s \subset \{v_{tt}, v_{tf}\}} m_{\mathcal{U}}^{D}(s) = m_{\mathcal{U}}^{D}(\{v_{tt}\}) + m_{\mathcal{U}}^{D}(\{v_{tt}, v_{tf}\}) = 0.9$ *and* $\sum_{s \subset s_{\mathcal{U}}(w)} m_{\mathcal{U}}^{D}(s) = \sum_{s \subset \{v_{tt}, v_{ft}\}} m_{\mathcal{U}}^{D}(s) = m_{\mathcal{U}}^{D}(\{v_{tt}\}) = 0.7$, *we have* $D \overset{0.9}{\models} r$ *and* $D \overset{0.7}{\models} w$ *(of course also* $D \overset{0.7}{\models} r$ *or* $D \overset{0.7}{\models} (r \wedge w)$, *but not* $D \overset{0.9}{\models} w$*).*

1.4 Proof theory

As I have said before, the simplicity of the semantics of classical propositional logic allows its great result: to find every correct consequence of a belief by analyzing only the syntactical level. This analysis is based on the notion of proof: a proposition is called syntactical consequence of a belief if it is provable from the believed premises. Consider a set of propositional symbols $\mathcal{U}$: in classical propositional logic a proof of $\psi \in \mathcal{L}_{\mathcal{U}}$ from a given belief (which can be at most one formula φ on $\mathcal{U}$) is a finite sequence $\psi_1, \ldots, \psi_m$ of formulas on $\mathcal{U}$ such that $\psi_m \equiv \psi$ and for every $k \in \{1, \ldots, m\}$ at least one of the following three conditions holds: ψ_k is an axiom, $\psi_k \equiv \varphi$ (if the formula φ is declared), or there are $i, j < k$ such that $\psi_j \equiv (\psi_i \rightarrow \psi_k)$. That is, a proof of classical propositional logic is based on axioms and premises and allows a sole inference rule, the so called "modus ponens": from φ and $(\varphi \rightarrow \psi)$ follows ψ. If such a proof exists, we say that ψ is provable from φ (in symbols $\varphi \vdash \psi$) if φ was declared, and that ψ is a theorem (in symbols $\vdash \psi$) if nothing was declared. We can generalize the classical notion of proof to the present theory by considering a finite sequence of graded formulas instead of simple formulas: the grade of the last one will be the grade of the syntactical consequence. Now we have to express axioms and premises as graded formulas and define the modus ponens for graded formulas. Axioms are certainties, so we can give them the grade 1. The task is more difficult for premises: given a support distribution $\{\langle \varphi_1, \alpha_1\rangle, \ldots, \langle \varphi_n, \alpha_n\rangle\}$, we cannot simply consider as premises the graded formulas $\langle \varphi_i, \alpha_i\rangle$, because different support distributions can describe the same support assignment; consider for instance the two equivalent support distributions $\{\langle \varphi, \alpha\rangle\}$ and $\left\{\left\langle (\varphi \wedge \varphi), \frac{\alpha}{2}\right\rangle, \left\langle \varphi, \frac{\alpha}{2}\right\rangle\right\}$: with such a definition from the first one we could infer $\langle \varphi, \alpha\rangle$, whereas from the second one only $\left\langle \varphi, \frac{\alpha}{2}\right\rangle$ or $\left\langle (\varphi \wedge \varphi), \frac{\alpha}{2}\right\rangle$ (which of course have the same meaning). We can overcome this problem if, given a support distribution $\{\langle \varphi_1, \alpha_1\rangle, \ldots, \langle \varphi_n, \alpha_n\rangle\}$, we consider as premises the graded formulas $\left\langle \bigvee \{\varphi_i \mid i \in I\}, \sum_{i \in I} \alpha_i \right\rangle$ where $I \subset \{1, \ldots, n\}$: in the previous example, with this definition (and considering as equal two graded formulas with the same meaning) from the first support distribution we can infer $\langle\varphi,\alpha\rangle$ and $\langle\bot,0\rangle$ (which has no influence, because 0 means "no support"), whereas from the second one $\langle\varphi,\alpha\rangle$, $\left\langle\varphi,\frac{\alpha}{2}\right\rangle$ (which is weaker than $\langle\varphi,\alpha\rangle$) and $\langle\bot,0\rangle$. The modus ponens can be extended in the following way: from $\langle\varphi,\alpha\rangle$ and $\langle(\varphi\to\psi),\beta\rangle$ infer $\langle\psi,\alpha+\beta-1\rangle$ (if $\alpha+\beta-1\geq 0$); that is, infer $\langle\psi,\max(\alpha+\beta-1,0)\rangle$ (since $\langle\psi,0\rangle$ means "no support to ψ"). This "graded modus ponens" generalizes the classical modus ponens (since $\alpha=\beta=1\Rightarrow\alpha+\beta-1=1$ and $\alpha=0\Rightarrow\alpha+\beta-1\leq 0$); and the choice of $\alpha+\beta-1$ (instead, for instance, of $\min(\alpha,\beta)$) can be justified by considering, for a graded formula $\langle\chi,\gamma\rangle$, $1-\gamma$ as the "lack of certainty" of χ : the lack of certainty of the conclusion of the graded modus ponens is the sum of the lacks of certainty of the premises (since $1-(\alpha+\beta-1)=(1-\alpha)+(1-\beta)$).

I introduce now the axioms of classical propositional logic (as graded formulas with grade 1) and give a precise definition of syntactical consequence. Of course there are many possible sets of axioms: I use the one proposed by Kleene (see [16]); for some axioms I also report the name with which they are known in the literature.

Definition 12 *Given a set of propositional symbols $\mathcal{U}$, $\mathcal{A}_{\mathcal{U}}\subset\mathcal{G}_{\mathcal{U}}$ is the set of the axioms on $\mathcal{U}$; i.e., the set of the graded formulas $\langle\varphi,1\rangle$ such that φ has one of the following forms (for suitable $\psi,\chi,\omega\in\mathcal{L}_{\mathcal{U}}$):*

$$\left.\begin{array}{ll}\bullet\ (\psi\to\top) & \end{array}\right\}\ \textit{definition of } \top$$

$$\left.\begin{array}{ll}\bullet\ (\bot\to\psi) & \end{array}\right\}\ \textit{definition of } \bot$$

$$\left.\begin{array}{ll}\bullet\ (\neg\neg\psi\to\psi) & \textit{(the law of contraposition)}\\ \bullet\ ((\psi\to\chi)\to((\psi\to\neg\chi)\to\neg\psi)) & \textit{(reductio ad absurdum)}\end{array}\right\}\ \textit{definition of } \neg$$

$$\left.\begin{array}{ll}\bullet\ (\psi\to(\psi\vee\chi)) & \\ \bullet\ (\chi\to(\psi\vee\chi)) & \\ \bullet\ ((\psi\to\omega)\to((\chi\to\omega)\to((\psi\vee\chi)\to\omega))) & \end{array}\right\}\ \textit{definition of } \vee$$

$$\left.\begin{array}{ll}\bullet\ ((\psi\wedge\chi)\to\psi) & \\ \bullet\ ((\psi\wedge\chi)\to\chi) & \\ \bullet\ (\psi\to(\chi\to(\psi\wedge\chi))) & \textit{(the law of adjunction)}\end{array}\right\}\ \textit{definition of } \wedge$$

$$\left.\begin{array}{ll}\bullet\ (\psi\to(\chi\to\psi)) & \textit{(the law of simplification)}\\ \bullet\ ((\psi\to\chi)\to((\psi\to(\chi\to\omega))\to(\psi\to\omega))) & \textit{(Frege's law)}\end{array}\right\}\ \textit{definition of } \to$$

$$\left.\begin{array}{ll}\bullet\ ((\psi\leftrightarrow\chi)\to(\psi\to\chi)) & \\ \bullet\ ((\psi\leftrightarrow\chi)\to(\chi\to\psi)) & \\ \bullet\ ((\psi\to\chi)\to((\chi\to\psi)\to(\psi\leftrightarrow\chi))) & \end{array}\right\}\ \textit{definition of } \leftrightarrow$$

Definition 13 *A formula ψ on $\mathcal{V}$ is a syntactical consequence with grade α of a support distribution $D=\{\langle\varphi_1,\alpha_1\rangle,\ldots,\langle\varphi_n,\alpha_n\rangle\}$ on $\mathcal{U}$ (in symbols $D\overset{\alpha}{\vdash}\psi$), if there is a finite sequence of graded formulas (called graded proof) $\langle\psi_1,\beta_1\rangle,\ldots,\langle\psi_m,\beta_m\rangle$ on $\mathcal{U}\cup\mathcal{V}$, such* *that* $\psi_m \equiv \psi$*,* $\beta_m \geq \alpha$*, and for every* $k \in \{1, \ldots, m\}$ *at least one of the following conditions holds:*

- $\langle \psi_k, \beta_k \rangle \in \mathcal{A}_{\mathcal{U} \cup \mathcal{V}}$*;*
- $\psi_k \equiv \bigvee \{\varphi_i \mid i \in I\}$ *and* $\beta_k = \sum_{i \in I} \alpha_i$ *for an* $I \subset \{1, \ldots, n\}$*;*
- *there are* $i, j < k$ *such that* $\psi_j \equiv (\psi_i \to \psi_k)$ *and* $\beta_k = \max\big(\beta_i + \beta_j - 1, 0\big)$*.*

Of course this notion of syntactical consequence generalizes the classical one: if D is a classical support distribution, then $D \vdash \psi \Leftrightarrow D \overset{1}{\vdash} \psi$.

Example 14 *Consider another time the situation of Example 5: we had a support distribution* $D = \{\langle (r \wedge w), 0.7 \rangle, \langle r, 0.2 \rangle\}$ *on* $\mathcal{U} \supset \{r, w\}$*; now I shall prove that* $D \overset{0.9}{\vdash} r$*. As first graded formula we take the premise* $\langle \bigvee \{(r \wedge w), r\}, 0.9 \rangle$*, which can represent, for instance,* $\langle ((\bot \vee (r \wedge w)) \vee r), 0.9 \rangle$*. Then we construct* $\langle (((\bot \vee (r \wedge w)) \vee r) \to r), 1 \rangle$ *using the axioms and the graded modus ponens:*

$$\begin{aligned}&\langle (\bot \to r), 1 \rangle, \langle ((r \wedge w) \to r), 1 \rangle, \\ &\langle ((\bot \to r) \to (((r \wedge w) \to r) \to ((\bot \vee (r \wedge w)) \to r))), 1 \rangle \in \mathcal{A}_{\mathcal{U}},\end{aligned}$$

since $1 + 1 - 1 = 1$*, from the first and the third of these axioms we obtain the graded formula* $\langle (((r \wedge w) \to r) \to ((\bot \vee (r \wedge w)) \to r)), 1 \rangle$*; and from this and the second axiom we obtain* $\langle ((\bot \vee (r \wedge w)) \to r), 1 \rangle$*. In the same way, from the axioms*

$$\begin{aligned}&\langle (r \to \top), 1 \rangle, \langle (r \to (\top \to r)), 1 \rangle, \\ &\langle ((r \to \top) \to ((r \to (\top \to r)) \to (r \to r))), 1 \rangle \in \mathcal{A}_{\mathcal{U}},\end{aligned}$$

follows the graded formula $\langle (r \to r), 1 \rangle$*. Finally, from these two conclusions and the axiom*

$$\langle (((\bot \vee (r \wedge w)) \to r) \to ((r \to r) \to (((\bot \vee (r \wedge w)) \vee r) \to r))), 1 \rangle \in \mathcal{A}_{\mathcal{U}},$$

we obtain (again using two times the graded modus ponens) the desired graded formula $\langle (((\bot \vee (r \wedge w)) \vee r) \to r), 1 \rangle$*. Now we can apply the graded modus ponens to this graded formula and the premise: we obtain* $\langle r, 0.9 \rangle$*; thus* $D \overset{0.9}{\vdash} r$*.*

1.5 Adequateness

The great result of classical propositional logic is that the proof theory is adequate: the semantical and the syntactical concepts of consequence are equivalent. This result consists of two parts: the easiest one is the soundness of the proof theory (that is, every syntactical consequence is a semantical one, or informally: every provable proposition is true, a basic requirement for a proof theory), whereas the difficult one is the completeness of the proof theory (that is, every semantical consequence is a syntactical one, or informally: every true proposition is provable, the aim of a proof theory). I shall now prove that this result extends to the present generalization; for this task I shall exploit the particular case of the completeness of classical propositional logic stated in Theorem 15: every tautology is a theorem, in symbols: $\models \psi \Rightarrow \vdash \psi$. For a proof (and a good presentation of classical propositional logic) see [11].

Theorem 15 *Given a set of propositional symbols $\mathcal{V}$ and a formula ψ on $\mathcal{V}$,*

$$\emptyset \overset{1}{\models} \psi \Rightarrow \emptyset \overset{1}{\vdash} \psi.$$

Lemma 16 (Soundness of the proof theory)

Given a support distribution D on $\mathcal{U}$, a formula ψ on $\mathcal{V}$ and a grade α,

$$D \overset{\alpha}{\vdash} \psi \Rightarrow D \overset{\alpha}{\models} \psi.$$

Proof. Let $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$; since $D \overset{\alpha}{\vdash} \psi$, there is a proof $\langle\psi_1, \beta_1\rangle, \ldots, \langle\psi_m, \beta_m\rangle \in \mathcal{G}_{\mathcal{U}\cup\mathcal{V}}$ of ψ with grade α.
I shall show by induction $D \overset{\beta_k}{\models} \psi_k$ for every $k \in \{1, \ldots, m\}$: this is enough, since for $k = m$ it means $D \overset{\beta_m}{\models} \psi$, i.e., $\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \beta_m \geq \alpha$, and so $D \overset{\alpha}{\models} \psi$.
Let $k \in \{1, \ldots, m\}$, to be proven is that $\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \beta_k$:
from the definition of graded proof we know that one of the following possibilities holds.

- $\langle\psi_k, \beta_k\rangle \in \mathcal{A}_{\mathcal{U}\cup\mathcal{V}}$; i.e., $\beta_k = 1$ and ψ_k has one of the 15 possible forms of Definition 12.
  It is easy to prove that in any case $s_{\mathcal{U}\cup\mathcal{V}}(\psi_k) = V_{\mathcal{U}\cup\mathcal{V}}$:
  for example, if $\psi_k \equiv (\chi \rightarrow \top)$ for a $\chi \in \mathcal{L}_{\mathcal{U}\cup\mathcal{V}}$, then (from Proposition 8)
  $s_{\mathcal{U}\cup\mathcal{V}}(\psi_k) = s_{\mathcal{U}\cup\mathcal{V}}((\chi \rightarrow \top)) = V_{\mathcal{U}\cup\mathcal{V}} \setminus (s_{\mathcal{U}\cup\mathcal{V}}(\chi) \setminus s_{\mathcal{U}\cup\mathcal{V}}(\top)) = V_{\mathcal{U}\cup\mathcal{V}}$,
  since $s_{\mathcal{U}\cup\mathcal{V}}(\top) = V_{\mathcal{U}\cup\mathcal{V}}$.
  Therefore $\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \sum\limits_{s \subset V_{\mathcal{U}\cup\mathcal{V}}} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = 1 \geq \beta_k$.

- $\psi_k \equiv \bigvee\{\varphi_i \mid i \in I\}$ and $\beta_k = \sum\limits_{i \in I} \alpha_i$ for an $I \subset \{1, \ldots, n\}$.
  Therefore (from Proposition 8) $\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \sum\limits_{s \subset \bigcup\limits_{i \in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)} m^D_{\mathcal{U}\cup\mathcal{V}}(s)$;

if $\bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) = V_{\mathcal{U}\cup\mathcal{V}}$, then $\sum_{s\subset \bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = 1 \geq \beta_k$, otherwise

$$\sum_{s\subset \bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \sum_{s\subset \bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)} \sum_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi j)=s} \alpha_j = \sum_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi j)\subset \bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)} \alpha_j \geq \sum_{i\in I} \alpha_i = \beta_k.$$

- There are $i, j < k$ such that $\psi_j \equiv (\psi_i \to \psi_k)$ and $\beta_k = \max\left(\beta_i + \beta_j - 1, 0\right)$.
  By induction $D \overset{\beta_i}{\models} \psi_i$ and $D \overset{\beta_j}{\models} \psi_j$;
  i.e., $\sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_i)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \beta_i$ and $\sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}\left(\psi_j\right)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \beta_j$.
  Consider the formula $(\psi_i \wedge \psi_j) \equiv (\psi_i \wedge (\psi_i \to \psi_k))$: from Proposition 8 follows that
  $$\begin{aligned} s_{\mathcal{U}\cup\mathcal{V}}(\psi_i) \cap s_{\mathcal{U}\cup\mathcal{V}}\left(\psi_j\right) &= s_{\mathcal{U}\cup\mathcal{V}}\left(\left(\psi_i \wedge \psi_j\right)\right) = s_{\mathcal{U}\cup\mathcal{V}}\left((\psi_i \wedge (\psi_i \to \psi_k))\right) = \\ &= s_{\mathcal{U}\cup\mathcal{V}}(\psi_i) \cap \left(V_{\mathcal{U}\cup\mathcal{V}} \setminus \left(s_{\mathcal{U}\cup\mathcal{V}}(\psi_i) \setminus s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)\right)\right) = \\ &= s_{\mathcal{U}\cup\mathcal{V}}(\psi_i) \cap s_{\mathcal{U}\cup\mathcal{V}}(\psi_k) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)\,. \end{aligned}$$
  Therefore
  $$\begin{aligned} \sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_k)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) &\geq \underbrace{\sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_i)\cap s_{\mathcal{U}\cup\mathcal{V}}\left(\psi_j\right)} m^D_{\mathcal{U}\cup\mathcal{V}}(s)}_{\geq 0} = \\ &= \underbrace{\sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_i)} m^D_{\mathcal{U}\cup\mathcal{V}}(s)}_{\geq \beta_i} + \underbrace{\sum_{s\subset s_{\mathcal{U}\cup\mathcal{V}}\left(\psi_j\right)} m^D_{\mathcal{U}\cup\mathcal{V}}(s)}_{\geq \beta_j} - \\ &\quad - \underbrace{\sum_{s:\ s\subset s_{\mathcal{U}\cup\mathcal{V}}(\psi_i) \text{ or } s\subset s_{\mathcal{U}\cup\mathcal{V}}\left(\psi_j\right)} m^D_{\mathcal{U}\cup\mathcal{V}}(s)}_{\leq 1} \geq \\ &\geq \max\left(\beta_i + \beta_j - 1, 0\right) = \beta_k. \ \blacksquare \end{aligned}$$

Lemma 17 (Completeness of the proof theory)

Given a support distribution D on $\mathcal{U}$, a formula ψ on $\mathcal{V}$ and a grade α,

$$D \overset{\alpha}{\models} \psi \Rightarrow D \overset{\alpha}{\vdash} \psi.$$

Proof. If $s_{\mathcal{U}\cup\mathcal{V}}(\psi) = V_{\mathcal{U}\cup\mathcal{V}}$, then, since $\sum_{s\subset V_{\mathcal{U}\cup\mathcal{V}}} m^{\emptyset}_{\mathcal{U}\cup\mathcal{V}}(s) = 1$, we have $\emptyset \overset{1}{\models} \psi$,
and from Theorem 15 follows $\emptyset \overset{1}{\vdash} \psi$;
i.e., there is a proof of ψ with grade 1 and without premises,
which of course is also a proof of ψ with grade α and premises D: therefore $D \overset{\alpha}{\vdash} \psi$.
Otherwise, if $s_{\mathcal{U}\cup\mathcal{V}}(\psi) \neq V_{\mathcal{U}\cup\mathcal{V}}$, let $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ and define
$I = \{i \mid s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)\}$, $\varphi \equiv \bigvee\{\varphi_i \mid i \in I\}$ and $\chi \equiv (\psi \wedge \neg\varphi)$.

From Proposition 8 follows: $s_{\mathcal{U}\cup\mathcal{V}}(\varphi) = \bigcup_{i\in I} s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)$
and $s_{\mathcal{U}\cup\mathcal{V}}(\chi) = s_{\mathcal{U}\cup\mathcal{V}}(\psi) \cap (V_{\mathcal{U}\cup\mathcal{V}} \setminus s_{\mathcal{U}\cup\mathcal{V}}(\varphi)) = s_{\mathcal{U}\cup\mathcal{V}}(\psi) \setminus s_{\mathcal{U}\cup\mathcal{V}}(\varphi)$,
therefore $s_{\mathcal{U}\cup\mathcal{V}}((\varphi \vee \chi)) = s_{\mathcal{U}\cup\mathcal{V}}(\varphi) \cup s_{\mathcal{U}\cup\mathcal{V}}(\chi) = s_{\mathcal{U}\cup\mathcal{V}}(\psi)$
and $s_{\mathcal{U}\cup\mathcal{V}}(((\varphi \vee \chi) \to \psi)) = V_{\mathcal{U}\cup\mathcal{V}} \setminus (s_{\mathcal{U}\cup\mathcal{V}}((\varphi \vee \chi)) \setminus s_{\mathcal{U}\cup\mathcal{V}}(\psi)) = V_{\mathcal{U}\cup\mathcal{V}}$.
As before we have $\emptyset \overset{1}{\models} ((\varphi \vee \chi) \to \psi)$, and so there is a proof $\langle \psi_1, \beta_1 \rangle, \ldots, \langle \psi_m, \beta_m \rangle \in \mathcal{G}_{\mathcal{U}\cup\mathcal{V}}$ of $((\varphi \vee \chi) \to \psi)$ with grade 1 and without premises: $\psi_m = ((\varphi \vee \chi) \to \psi)$ and $\beta_m = 1$. Considering the premises D, we can add to this proof the following sequence of graded formulas:

$$\left\langle \varphi, \sum_{i\in I} \alpha_i \right\rangle, \langle (\varphi \to (\varphi \vee \chi)), 1 \rangle, \left\langle (\varphi \vee \chi), \sum_{i\in I} \alpha_i \right\rangle, \left\langle \psi, \sum_{i\in I} \alpha_i \right\rangle,$$

where the first one is a premise, the second one an axiom
and the last two are the results of two applications of the graded modus ponens.
Since $\sum_{i\in I} \alpha_i = \sum_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} \alpha_i = \sum_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} \sum_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) = s} \alpha_i = \sum_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \alpha$,
the whole sequence of graded formulas is a proof of ψ with grade α and premises D; therefore $D \overset{\alpha}{\vdash} \psi$. ■

Theorem 18 (ADEQUATENESS OF THE PROOF THEORY)

Given a support distribution D on $\mathcal{U}$, a formula ψ on $\mathcal{V}$ and a grade α,

$$D \overset{\alpha}{\vdash} \psi \Leftrightarrow D \overset{\alpha}{\models} \psi.$$

Proof. Directly from Lemmas 16 and 17. ■

1.6 Conclusions and consistency

In the previous section we have seen that the semantical and the syntactical concepts of consequence are equivalent, so I shall now call these consequences the (inductive or deductive) conclusions of a support distribution.

Definition 19 *The (inductive) conclusions with grade α on $\mathcal{V}$ of a support distribution D on $\mathcal{U}$ are the elements of*

$$\mathcal{C}^{\alpha}_{\mathcal{V}}(D) = \left\{ \psi \in \mathcal{L}_{\mathcal{V}} \mid D \overset{\alpha}{\models} \psi \right\} = \left\{ \psi \in \mathcal{L}_{\mathcal{V}} \mid D \overset{\alpha}{\vdash} \psi \right\}.$$

Of course it is sensible to accept as inductive conclusions only those with a grade close to 1, whereas the deductive conclusions are the extreme case with grade 1.

Definition 20 *The deductive conclusions on $\mathcal{V}$ of a support distribution D on $\mathcal{U}$ are the elements of*

$$\mathcal{D}_{\mathcal{V}}(D) = \mathcal{C}^1_{\mathcal{V}}(D).$$

If the support distribution D is classical, the deductive conclusions are the conclusions of classical propositional logic: $\mathcal{D}_{\mathcal{V}}(D) = \{\psi \in \mathcal{L}_{\mathcal{V}} \mid D \models \psi\} = \{\psi \in \mathcal{L}_{\mathcal{V}} \mid D \vdash \psi\}$. In particular, if $D = \emptyset$ the deductive conclusions are the tautologies of classical propositional logic.

Definition 21 *The tautologies on $\mathcal{V}$ are the elements of*

$$\mathcal{T}_{\mathcal{V}} = \mathcal{D}_{\mathcal{V}}(\emptyset).$$

From the definition of semantical consequence follows directly that for any support distribution D and for any grades $\alpha \geq \beta$, $\mathcal{D}_{\mathcal{V}}(D) \subset \mathcal{C}^{\alpha}_{\mathcal{V}}(D) \subset \mathcal{C}^{\beta}_{\mathcal{V}}(D) \subset \mathcal{C}^{0}_{\mathcal{V}}(D) = \mathcal{L}_{\mathcal{V}}$: the deductive conclusions are inductive ones for every grade, and every proposition is a trivial (i.e., with grade 0) conclusion. From the definition of syntactical consequence follows directly that for any support distribution D, $\mathcal{T}_{\mathcal{V}} \subset \mathcal{D}_{\mathcal{V}}(D)$: the tautologies are deductive conclusions of every possible belief; therefore from the definition of $\mathcal{T}_{\mathcal{V}}$ follows $\mathcal{T}_{\mathcal{V}} = \bigcap_{D} \mathcal{D}_{\mathcal{V}}(D)$ (where D can be any support distribution): the tautologies and nothing more are what we are compelled to believe. Considering the particular role of the tautologies and noticing that $\mathcal{V} \subset \mathcal{W} \Rightarrow \mathcal{C}^{\alpha}_{\mathcal{V}}(D) = \mathcal{C}^{\alpha}_{\mathcal{W}}(D) \cap \mathcal{L}_{\mathcal{V}}$ (for every grade α and support distribution D; thus in particular: $\mathcal{T}_{\mathcal{V}} = \mathcal{T}_{\mathcal{W}} \cap \mathcal{L}_{\mathcal{V}}$) the following definitions are sensible.

Definition 22 *Given two formulas φ and ψ on $\mathcal{V}$, φ and ψ are equivalent (in symbols $\varphi \sim \psi$) if $(\varphi \leftrightarrow \psi) \in \mathcal{T}_{\mathcal{V}}$.*

Definition 23 *Given two formulas φ and ψ on $\mathcal{V}$, φ implies ψ (in symbols $\varphi \rightsquigarrow \psi$) if $(\varphi \to \psi) \in \mathcal{T}_{\mathcal{V}}$.*

Proposition 24 *Given two formulas φ and ψ on $\mathcal{V}$,*

- $\varphi \sim \psi \Leftrightarrow s_{\mathcal{V}}(\varphi) = s_{\mathcal{V}}(\psi)$ *(φ and ψ are equivalent if and only if they have the same models, and therefore $\sim$ is an equivalence relation on $\mathcal{L}_{\mathcal{V}}$),*

- $\varphi \rightsquigarrow \psi \Leftrightarrow s_{\mathcal{V}}(\varphi) \subset s_{\mathcal{V}}(\psi)$ *(φ implies ψ if and only if every model of φ is a model of ψ).*

Proof. Using Proposition 8:

- $\varphi \sim \psi \Leftrightarrow (\varphi \leftrightarrow \psi) \in \mathcal{T}_{\mathcal{V}} \Leftrightarrow \emptyset \overset{1}{\models} (\varphi \leftrightarrow \psi) \Leftrightarrow \sum_{s \subset s_{\mathcal{V}}((\varphi \leftrightarrow \psi))} m_{\mathcal{V}}^{\emptyset}(s) = 1 \Leftrightarrow$
$\Leftrightarrow s_{\mathcal{V}}((\varphi \leftrightarrow \psi)) = V_{\mathcal{V}} \Leftrightarrow s_{\mathcal{V}}(\varphi) = s_{\mathcal{V}}(\psi)$;
- $\varphi \rightsquigarrow \psi \Leftrightarrow (\varphi \to \psi) \in \mathcal{T}_{\mathcal{V}} \Leftrightarrow \emptyset \overset{1}{\models} (\varphi \to \psi) \Leftrightarrow \sum_{s \subset s_{\mathcal{V}}((\varphi \to \psi))} m_{\mathcal{V}}^{\emptyset}(s) = 1 \Leftrightarrow$
$\Leftrightarrow s_{\mathcal{V}}((\varphi \to \psi)) = V_{\mathcal{V}} \Leftrightarrow s_{\mathcal{V}}(\varphi) \subset s_{\mathcal{V}}(\psi)$. ■

The fact that a proposition is a conclusion does not imply that its negation is not supported; thus it can be interesting to give prominence to the conclusions whose negations do not have any support (I shall call these strong conclusions).

Definition 25 *The strong conclusions with grade α on $\mathcal{V}$ of a support distribution D on $\mathcal{U}$ are the elements of*

$$\mathcal{S}_{\mathcal{V}}^{\alpha}(D) = \left\{ \psi \in \mathcal{L}_{\mathcal{V}} \mid \psi \in \mathcal{C}_{\mathcal{V}}^{\alpha}(D),\ \forall \beta > 0\ \ \neg\psi \notin \mathcal{C}_{\mathcal{V}}^{\beta}(D) \right\}.$$

Of course for any grade α we have $\mathcal{S}_{\mathcal{V}}^{\alpha}(D) \subset \mathcal{C}_{\mathcal{V}}^{\alpha}(D)$, but for the grade $\alpha = 1$ it seems natural to me to expect $\mathcal{S}_{\mathcal{V}}^{1}(D) = \mathcal{D}_{\mathcal{V}}(D)$ (that is, every deductive conclusion is a strong one): in general this is false, we have only $\mathcal{S}_{\mathcal{V}}^{1}(D) \subset \mathcal{D}_{\mathcal{V}}(D)$ (there can be deductive conclusions that are not strong); and even worse: in general $\mathcal{T}_{\mathcal{V}} \subset \mathcal{S}_{\mathcal{V}}^{1}(D)$ is false (even tautologies do not have to be strong). This is due to a problem (perhaps the only one) of classical propositional logic: the possibility of inconsistency. Informally, a set of statements is called inconsistent if some of its elements are contradictory (i.e., their conjunction is impossible). In classical propositional logic from an impossible (i.e., without models) proposition we can conclude everything, so we say that a belief is inconsistent if every proposition is a conclusion of it: Proposition 26 states that this is the case if and only if from the belief we can conclude the falsehood. In the classical case the conclusions are certainties, thus the inconsistency can only be complete: we are sure of the falsehood; on the contrary, in the present theory the inconsistency can be partial: we believe the falsehood with a grade $\alpha \in (0, 1)$.

Proposition 26 *Given a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$ and a grade α,*

$$\mathcal{C}_{\mathcal{V}}^{\alpha}(D) = \mathcal{L}_{\mathcal{V}} \ \Leftrightarrow\ D \overset{\alpha}{\models} \bot \ \Leftrightarrow \sum_{\varphi_i \sim \bot} \alpha_i \geq \alpha.$$

Proof. Trivially, since $\perp \in \mathcal{L}_\mathcal{V}$, $\mathcal{C}^\alpha_\mathcal{V}(D) = \mathcal{L}_\mathcal{V} \Rightarrow D \overset{\alpha}{\models} \perp$.
On the other hand, $D \overset{\alpha}{\models} \perp \Rightarrow D \overset{\alpha}{\vdash} \perp$, and $(\perp \to \psi)$ is an axiom for every $\psi \in \mathcal{L}_\mathcal{V}$, so by applying the graded modus ponens we obtain $D \overset{\alpha}{\vdash} \psi$ for every $\psi \in \mathcal{L}_\mathcal{V}$;
i.e., $\mathcal{C}^\alpha_\mathcal{V}(D) = \mathcal{L}_\mathcal{V}$.
For the second equivalence:
$D \overset{\alpha}{\models} \perp \Leftrightarrow \sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\perp)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \geq \alpha \overset{s_{\mathcal{U}\cup\mathcal{V}}(\perp)=\emptyset}{\Leftrightarrow} m^D_{\mathcal{U}\cup\mathcal{V}}(\emptyset) \geq \alpha \Leftrightarrow$
$\Leftrightarrow \sum\limits_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i)=\emptyset} \alpha_i \geq \alpha \Leftrightarrow \sum\limits_{\varphi_i \sim \perp} \alpha_i \geq \alpha.$ ■

Definition 27 *Given a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$, the (degree of) inconsistency of D is $I_D = \sum\limits_{\varphi_i \sim \perp} \alpha_i = m^D_\mathcal{U}(\emptyset)$; if $I_D = 0$, then D is called consistent, if $I_D \in (0,1)$, then D is called partially inconsistent, and if $I_D = 1$, then D is called completely inconsistent.*

Given a support distribution D on $\mathcal{U}$, if D is consistent, then $\mathcal{S}^1_\mathcal{V}(D) = \mathcal{D}_\mathcal{V}(D)$ (every deductive conclusion is a strong one), if D is inconsistent, then, for every grade α, $\mathcal{S}^\alpha_\mathcal{V}(D) = \emptyset$ (there are not strong conclusions of any grade), and if D is completely inconsistent, then $\mathcal{D}_\mathcal{V}(D) = \mathcal{L}_\mathcal{V}$ (every proposition is a deductive conclusion).

Of course we would like to avoid inconsistency (because concluding the falsehood, even with a very small grade, is not a form of correct reasoning): if a support distribution is inconsistent, then we have to transform it into a consistent one, which should be as close as possible to the original one. Given an inconsistent support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$, the easiest way to transform it is to eliminate the supports assigned to the formulas which are at the origin of the inconsistency; i.e., to consider the new (consistent) support distribution $D' = \{\langle\varphi_1, \beta_1\rangle, \ldots, \langle\varphi_n, \beta_n\rangle\}$, where $\beta_i = 0$ if $\varphi_i \sim \perp$, and $\beta_i = \alpha_i$ otherwise: from the standpoint of the associated SSD this means considering the new SSD $m^{D'}_\mathcal{U}$ such that $m^{D'}_\mathcal{U}(\emptyset) = 0$, $m^{D'}_\mathcal{U}(V_\mathcal{U}) = m^D_\mathcal{U}(V_\mathcal{U}) + m^D_\mathcal{U}(\emptyset)$ and otherwise $m^{D'}_\mathcal{U} = m^D_\mathcal{U}$. This method is very simple, but it leads to debatable consequences:

$$\mathcal{C}^\alpha_\mathcal{V}(D') = \begin{cases} \mathcal{C}^{\alpha+I_D}_\mathcal{V}(D) & \text{if } \alpha \in [0, 1-I_D], \\ \mathcal{T}_\mathcal{V} & \text{if } \alpha \in [1-I_D, 1]; \end{cases}$$

that is, the grade of every conclusion is reduced by I_D, independently of how close to 1 it was, and in particular there are no more deductive conclusions (apart, obviously, from the tautologies): the problem of this method is that it deals with every conclusion (inductive or deductive) in the same way, whereas I think that deductive conclusions should be considered as a very particular case. If the inconsistency is partial, deductive conclusions should be maintained; this can be done by eliminating the supports assigned to the formulas which are at the origin of the inconsistency, and renormalizing to 1 the remaining supports; i.e., by considering the new (consistent) support distribution $\overline{D} = \overline{\{\langle\varphi_1,\alpha_1\rangle,\ldots,\langle\varphi_n,\alpha_n\rangle\}} = \{\langle\varphi_1,\beta_1\rangle,\ldots,\langle\varphi_n,\beta_n\rangle\}$, where $\beta_i = 0$ if $\varphi_i \sim \bot$, and $\beta_i = \frac{\alpha_i}{1-I_D}$ otherwise. This renormalization reminds of a probabilistic way to behave: in fact from the standpoint of the associated SSD this means considering the new SSD $m_{\mathcal{U}}^{\overline{D}}$ such that $m_{\mathcal{U}}^{\overline{D}}(\emptyset) = 0$ and otherwise $m_{\mathcal{U}}^{\overline{D}}(s) = \frac{m_{\mathcal{U}}^{D}(s)}{1-m_{\mathcal{U}}^{D}(\emptyset)}$; if we consider (formally) a SSD as a probability distribution on $\mathcal{P}(V_{\mathcal{U}})$, this process is the conditioning on $\mathcal{P}(V_{\mathcal{U}}) \setminus \{\emptyset\}$. With this method (which I shall call renormalization):

$$\mathcal{C}_{\mathcal{V}}^{\alpha}\left(\overline{D}\right) = \mathcal{C}_{\mathcal{V}}^{\alpha(1-I_D)+I_D}(D);$$

that is, the grade of every conclusion is reduced by a number in $[0, I_D]$, depending on how close to 1 it was (the closer, the less reduced), and in particular the deductive conclusions are maintained and become strong: $\mathcal{S}_{\mathcal{V}}^{1}\left(\overline{D}\right) = \mathcal{D}_{\mathcal{V}}\left(\overline{D}\right) = \mathcal{D}_{\mathcal{V}}(D)$.

If the inconsistency is complete (i.e., $m_{\mathcal{U}}^{D}(\emptyset) = 1$), the first method transforms the SSD into the one associated with the empty support distribution (i.e., $m_{\mathcal{U}}^{D'} = m_{\mathcal{U}}^{\emptyset}$), whereas the renormalization is inapplicable. A complete inconsistency means the certainty of the falsehood (that is, the falsehood is a deductive conclusion): the first method is so strong and blind to consider this case like any other (in this method inductive or deductive conclusions are not distinguished), whereas the second considers this case as insurmountable and stops. In this case too, I prefer the renormalization: the certainty of the falsehood is the only situation from which we cannot get out, because we cannot clutch to any certainty, since no certainty is reliable any more (not even the tautologies: they are not surer than the falsehood); in such a situation a computer program is expected to crash. Beyond the metaphors, I think that it is a good thing to stop in such a situation (remembering that we are modelling and thus simplifying the human reasoning), and in the following I shall use the renormalization method (even if hundreds of sensible methods could be proposed). Anyway, the problem of inconsistency will return only when I shall analyze the dynamic aspect of the theory (the evolution of a belief): until this moment "support distribution" will always mean "consistent support distribution".

The conclusions of classical propositional logic are closed under inference: this means that if φ and $(\varphi \to \psi)$ are conclusions, then ψ too (in particular, if φ and ψ are conclusions, then $(\varphi \wedge \psi)$ too, since $(\varphi \to (\psi \to (\varphi \wedge \psi)))$ is an axiom). This property is very elegant and practical, so that intuitively it seems to be necessary; a necessity which was postulated by Hempel in [13] with regard to the theorems which are accepted in science: the set of all accepted statements should be consistent and closed under inference[9]. The

[9]More precisely, Hempel formulated three necessary conditions of rationality for the set K of all statements accepted as presumably true by scientists at time t (but the third one is not interesting in the actual context):

- "Any logical consequence of a set of accepted statements is likewise an accepted statement; or, K contains all logical consequences of any of its subclasses."

requirement of consistency is indisputable and has already been discussed, whereas the second requirement is problematic: the deductive conclusions of a support distribution fulfill it (because the graded modus ponens reduces to the standard modus ponens if the grades are both 1), but in general the inductive ones do not. This could seem a big problem, surely it is a very important point: here we can see very clearly the dichotomy in the formalizations of inductive inferences mentioned in the introduction (Section 1.1), between the formalizations which merely extend the notion of logical consequence (accepting only one level of conclusions) and the ones which do this by generalizing the notion of knowledge (allowing many levels of conclusions). The formalizations of the first kind follow Hempel's postulate: the set of conclusions must always be closed under inference (see for instance [2]); but this leads in general to big problems, because, given a set of believed statements S, accepting a new proposition s compels us to accept every statement in the closure of the set $S \cup \{s\}$, and given some new propositions $s_1, \ldots, s_n$ that we would like to accept, it can be that the closures of the n sets $S \cup \{s_i\}$ are incompatible (i.e., their union is inconsistent). The formalizations of the latter kind reject Hempel's postulate because of the so called "Kyburg's lottery paradox" (see [18]): consider a lottery with n tickets, exactly one of which gives the prize; in a game of chance it is sensible to accept the probability as belief, thus the belief in every one of the n propositions "The i-th ticket does not win" is $\frac{n-1}{n}$, which can be made as high as one likes by increasing n, and so (for an n large enough) we have to accept all these propositions, but their conjunction "Anyone of the tickets wins" is in contradiction with the premises. For this reason, Kyburg has modified Hempel's postulate in the following way: if we accept a statement, then we have to accept its tautological consequences and we cannot accept its negation (for a comprehensive discussion of this topic, see for instance [21]). Proposition 28 shows that the inductive conclusions of the present theory fulfill this postulate if the grade is bigger than $\frac{1}{2}$ (since it is sensible to accept as inductive conclusions only those with a grade close to 1, this is not a restriction) and the inductive strong conclusions fulfill it for any positive grade.

Proposition 28 *Given a support distribution D on $\mathcal{U}$ and two formulas φ and ψ on $\mathcal{V}$,*

- *if $\varphi \in \mathcal{C}^{\alpha}_{\mathcal{V}}(D)$, then $\varphi \rightsquigarrow \psi \Rightarrow \psi \in \mathcal{C}^{\alpha}_{\mathcal{V}}(D)$ and (if $\alpha > \frac{1}{2}$) $\neg\varphi \notin \mathcal{C}^{\alpha}_{\mathcal{V}}(D)$;*
- *if $\varphi \in \mathcal{S}^{\alpha}_{\mathcal{V}}(D)$, then $\varphi \rightsquigarrow \psi \Rightarrow \psi \in \mathcal{S}^{\alpha}_{\mathcal{V}}(D)$ and (if $\alpha > 0$) $\neg\varphi \notin \mathcal{S}^{\alpha}_{\mathcal{V}}(D)$.*

Proof. If $\varphi \in \mathcal{C}^{\alpha}_{\mathcal{V}}(D)$, then (from Proposition 24)

$$\varphi \rightsquigarrow \psi \Rightarrow s_{\mathcal{U}\cup\mathcal{V}}(\varphi) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi) \Rightarrow \sum_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^{D}_{\mathcal{U}\cup\mathcal{V}}(s) \geq \sum_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\varphi)} m^{D}_{\mathcal{U}\cup\mathcal{V}}(s) \geq \alpha \Rightarrow$$
$$\Rightarrow \psi \in \mathcal{C}^{\alpha}_{\mathcal{V}}(D);$$

- "The set K of accepted statements is logically consistent."

if $\alpha > \frac{1}{2}$, then (from Proposition 8, since $m^D_{\mathcal{U}\cup\mathcal{V}}(\emptyset) = 0$)
$\sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\neg\varphi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \leq 1 - \sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\varphi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \leq 1 - \alpha < \frac{1}{2} \Rightarrow \neg\varphi \notin \mathcal{C}^\alpha_\mathcal{V}(D)$.
If $\varphi \in \mathcal{S}^\alpha_\mathcal{V}(D)$, then (from Propositions 8 and 24, since $\psi \in \mathcal{C}^\alpha_\mathcal{V}(D)$)
$\varphi \rightsquigarrow \psi \Rightarrow s_{\mathcal{U}\cup\mathcal{V}}(\varphi) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi) \Rightarrow s_{\mathcal{U}\cup\mathcal{V}}(\neg\psi) \subset s_{\mathcal{U}\cup\mathcal{V}}(\neg\varphi) \Rightarrow$
$\Rightarrow \sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\neg\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) \leq \sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\neg\varphi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = 0 \Rightarrow \psi \in \mathcal{S}^\alpha_\mathcal{V}(D)$;
if $\alpha > 0$, then (from the definition of $\mathcal{S}^\alpha_\mathcal{V}(D)$) $\neg\varphi \notin \mathcal{C}^\alpha_\mathcal{V}(D) \Rightarrow \neg\varphi \notin \mathcal{S}^\alpha_\mathcal{V}(D)$. ■

1.7 Beliefs and plausibilities

I have introduced the support distribution as a description of the belief of a person on a topic, and then I have studied what conclusions this person is consequently compelled to believe (at a certain grade) if he/she reasons correctly. Thus, assuming the correctness of reasoning, we can describe the belief of this person better by assigning to every proposition the maximum grade with which this proposition is believed.

Definition 29 *Given a support distribution D on $\mathcal{U}$, the associated belief on $\mathcal{V}$ is the function*

$$\begin{array}{rccl} b^D_\mathcal{V}: & \mathcal{L}_\mathcal{V} & \longrightarrow & [0,1] \\ & \varphi & \longmapsto & \max\{\alpha \in [0,1] \mid \varphi \in \mathcal{C}^\alpha_\mathcal{V}(D)\}. \end{array}$$

Proposition 30 shows that the belief in a non-tautological proposition is the sum of the supports assigned to propositions from which it is implied.

Proposition 30 *Given a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$ and a formula ψ on $\mathcal{V}$,*

$$b^D_\mathcal{V}(\psi) = \sum_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \begin{cases} \sum\limits_{\varphi_i \rightsquigarrow \psi} \alpha_i & \textit{if } \psi \notin \mathcal{T}_\mathcal{V}, \\ 1 & \textit{if } \psi \in \mathcal{T}_\mathcal{V}. \end{cases}$$

Proof. The first equivalence follows directly from the definition of semantical consequence. For the second one:
if $\psi \in \mathcal{T}_\mathcal{V}$, then $b^D_\mathcal{V}(\psi) = 1$, since $\mathcal{T}_\mathcal{V} \subset \mathcal{D}_\mathcal{V}(D) = \mathcal{C}^1_\mathcal{V}(D)$;
if $\psi \notin \mathcal{T}_\mathcal{V}$, then $\psi \notin \mathcal{D}_\mathcal{V}(\emptyset) \Rightarrow \sum\limits_{s \subset s_\mathcal{V}(\psi)} m^\emptyset_\mathcal{V}(s) < 1 \Rightarrow s_\mathcal{V}(\psi) \neq V_\mathcal{V}$,
therefore from the definition of SSD and from Proposition 24 follows
$b^D_\mathcal{V}(\psi) = \sum\limits_{s \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \sum\limits_{s_{\mathcal{U}\cup\mathcal{V}}(\varphi_i) \subset s_{\mathcal{U}\cup\mathcal{V}}(\psi)} \alpha_i = \sum\limits_{\varphi_i \rightsquigarrow \psi} \alpha_i$. ■

Lemma 31 will be very useful in the following proofs.

Lemma 31 *Given a finite set A,*

$$\sum_{B\subset A}(-1)^{|B|}=\delta_0^{|A|}.$$

Proof. If $|A|=0$, then $A=\emptyset\Rightarrow\sum_{B\subset A}(-1)^{|B|}=(-1)^{|\emptyset|}=1$.
If $|A|\neq 0$, then
$\sum_{B\subset A}(-1)^{|B|}=\sum_{n=0}^{|A|}\sum_{B\subset A,\ |B|=n}(-1)^{|B|}=\sum_{n=0}^{|A|}\binom{|A|}{n}(-1)^n=(1+(-1))^{|A|}=0.$ ■

Theorem 32 *Given a support distribution D on $\mathcal{U}$, the associated belief on $\mathcal{V}$ has the following properties:*

- *given two formulas φ and ψ on $\mathcal{V}$, $\varphi\sim\psi\Rightarrow b_{\mathcal{V}}^{D}(\varphi)=b_{\mathcal{V}}^{D}(\psi)$;*
- *$b_{\mathcal{V}}^{D}(\bot)=0$;*
- *$b_{\mathcal{V}}^{D}(\top)=1$;*
- *given a finite nonempty set F of formulas on $\mathcal{V}$,*

$$b_{\mathcal{V}}^{D}\left(\bigvee F\right)\geq\sum_{\emptyset\neq F'\subset F}(-1)^{|F'|+1}b_{\mathcal{V}}^{D}\left(\bigwedge F'\right).$$

Proof. The first three properties follow easily from Propositions 30, 24 and 8.
The fourth one is more difficult, but defining $F(s)=\{\varphi\in F\mid s\subset s_{\mathcal{U}\cup\mathcal{V}}(\varphi)\}$ for $s\subset V_{\mathcal{U}\cup\mathcal{V}}$, and considering Propositions 30 and 8 and Lemma 31, we obtain:

$$\begin{aligned}
\sum_{\emptyset\neq F'\subset F}(-1)^{|F'|+1}b_{\mathcal{V}}^{D}\left(\bigwedge F'\right) &= \sum_{\emptyset\neq F'\subset F}(-1)^{|F'|+1}\sum_{s\subset\bigcap_{\varphi\in F'}s_{\mathcal{U}\cup\mathcal{V}}(\varphi)}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)=\\
&=\sum_{s\subset V_{\mathcal{U}\cup\mathcal{V}}}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)\sum_{\emptyset\neq F'\subset F(s)}(-1)^{|F'|+1}=\\
&=\sum_{s\subset V_{\mathcal{U}\cup\mathcal{V}}}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)(-1)\left(\delta_0^{|F(s)|}-(-1)^{|\emptyset|}\right)=\\
&=\sum_{s\subset V_{\mathcal{U}\cup\mathcal{V}}}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)\left(1-\delta_0^{|F(s)|}\right)=\sum_{s\subset V_{\mathcal{U}\cup\mathcal{V}},\ F(s)\neq\emptyset}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)\leq\\
&\leq\sum_{s\subset\bigcup_{\varphi\in F}s_{\mathcal{U}\cup\mathcal{V}}(\varphi)}m_{\mathcal{U}\cup\mathcal{V}}^{D}(s)=b_{\mathcal{V}}^{D}\left(\bigvee F\right).\ \blacksquare
\end{aligned}$$

Therefore, the present theory states that a person who reasons correctly believes two equivalent propositions with the same strength, does not believe in the falsehood, is sure of the truth and believes in the disjunction of propositions in agreement with the fourth property of Theorem 32, which for two and three propositions, respectively, states the following ($\varphi, \psi, \chi \in \mathcal{L}_\mathcal{V}$):

$$b^D_\mathcal{V}((\varphi \vee \psi)) \geq b^D_\mathcal{V}(\varphi) + b^D_\mathcal{V}(\psi) - b^D_\mathcal{V}((\varphi \wedge \psi));$$

$$\begin{aligned} b^D_\mathcal{V}(((\varphi \vee \psi) \vee \chi)) \geq\ & b^D_\mathcal{V}(\varphi) + b^D_\mathcal{V}(\psi) + b^D_\mathcal{V}(\chi) - \\ & - b^D_\mathcal{V}((\varphi \wedge \psi)) - b^D_\mathcal{V}((\varphi \wedge \chi)) - b^D_\mathcal{V}((\psi \wedge \chi)) + b^D_\mathcal{V}(((\varphi \wedge \psi) \wedge \chi)). \end{aligned}$$

Let $B_\mathcal{V}$ be the set of the functions $b \in [0,1]^{\mathcal{L}_\mathcal{V}}$ which satisfy the four properties of Theorem 32. Theorem 33 states that these properties are sufficient to characterize an acceptable belief, in the sense that for every function $b \in B_\mathcal{V}$, there is a support distribution on $\mathcal{V}$ whose associated belief is b. Therefore it is sensible to define a belief on $\mathcal{V}$ as an element of $B_\mathcal{V}$.

Theorem 33 *Given a function $b \in B_\mathcal{V}$, there is a support distribution D on $\mathcal{V}$ such that $b^D_\mathcal{V} = b$.*

Proof. For every $s \subset V_\mathcal{V}$ choose a $\varphi_s \in \mathcal{L}_\mathcal{V}$ such that $s_\mathcal{V}(\varphi_s) = s$
(the existence is assured by Proposition 7) and define $\alpha_s = \sum_{s' \subset s} (-1)^{|s \setminus s'|} b(\varphi_{s'})$.
I shall prove (using repeatedly Propositions 8 and 24)
that $D = \{\langle \varphi_s, \alpha_s \rangle \mid s \subset V_\mathcal{V}\}$ is a support distribution on $\mathcal{V}$ and that $b^D_\mathcal{V} = b$.
Consider an $s \subset V_\mathcal{V}$: if $s = \emptyset$ then $\alpha_s = b(\varphi_\emptyset) = b(\bot) = 0$;
otherwise $s = \{v_1, \ldots, v_n\}$: defining $F = \left\{\varphi_{s \setminus \{v_i\}} \mid i \in \{1, \ldots, n\}\right\}$ we obtain
$s_\mathcal{V}(\bigvee F) = \bigcup_{i=1}^{n} s_\mathcal{V}\left(\varphi_{s \setminus \{v_i\}}\right) = \bigcup_{i=1}^{n} s \setminus \{v_i\} = s$,
whereas if $s' \subsetneq s$, defining $F' = \left\{\varphi_{s \setminus \{v_i\}} \mid v_i \in s \setminus s'\right\}$ we obtain
$s_\mathcal{V}(\bigwedge F') = \bigcap_{i:\ v_i \in s \setminus s'} s_\mathcal{V}\left(\varphi_{s \setminus \{v_i\}}\right) = \bigcap_{i:\ v_i \in s \setminus s'} s \setminus \{v_i\} = s'$,
therefore $\alpha_s = b(\varphi_s) + \sum_{s' \subsetneq s} (-1)^{|s \setminus s'|} b(\varphi_{s'}) = b(\bigvee F) + \sum_{\emptyset \neq F' \subset F} (-1)^{|F'|} b(\bigwedge F') \geq 0$.
Since (using Lemma 31)

$$\begin{aligned} \sum_{s \subset V_\mathcal{V}} \alpha_s &= \sum_{s \subset V_\mathcal{V}} \sum_{s' \subset s} (-1)^{|s \setminus s'|} b(\varphi_{s'}) = \sum_{s' \subset V_\mathcal{V}} b(\varphi_{s'}) \sum_{s' \subset s \subset V_\mathcal{V}} (-1)^{|s \setminus s'|} = \\ &= \sum_{s' \subset V_\mathcal{V}} b(\varphi_{s'}) \delta_0^{|V_\mathcal{V} \setminus s'|} = b\left(\varphi_{V_\mathcal{V}}\right) = b(\top) = 1 \end{aligned}$$

and $\sum_{s \subset V_\mathcal{V}:\ \varphi_s \sim \bot} \alpha_s = \alpha_\emptyset = 0$,
D is a support distribution on $\mathcal{V}$ (because of course $s \neq s' \Rightarrow \varphi_s \neq \varphi_{s'}$).
To prove $b^D_\mathcal{V} = b$, consider $\psi \in \mathcal{L}_\mathcal{V}$: from Proposition 30 it follows that
if $\psi \notin \mathcal{T}_\mathcal{V}$, then (using Lemma 31)

$$\begin{aligned} b^D_\mathcal{V}(\psi) &= \sum_{s \subset V_\mathcal{V}:\ \varphi_s \rightsquigarrow \psi} \alpha_s = \sum_{s \subset s_\mathcal{V}(\psi)} \sum_{s' \subset s} (-1)^{|s \setminus s'|} b(\varphi_{s'}) = \sum_{s' \subset s_\mathcal{V}(\psi)} b(\varphi_{s'}) \sum_{s' \subset s \subset s_\mathcal{V}(\psi)} (-1)^{|s \setminus s'|} = \\ &= \sum_{s' \subset s_\mathcal{V}(\psi)} b(\varphi_{s'}) \delta_0^{|s_\mathcal{V}(\psi) \setminus s'|} = b\left(\varphi_{s_\mathcal{V}(\psi)}\right) = b(\psi), \end{aligned}$$

and if $\psi \in \mathcal{T}_{\mathcal{V}}$, then $s_{\mathcal{V}}(\psi) = V_{\mathcal{V}} \Rightarrow b(\psi) = b(\top) = 1 = b^D_{\mathcal{V}}(\psi)$;
therefore $b^D_{\mathcal{V}} = b$. ■

Of course, given a function $b \in B_{\mathcal{V}}$, the support distribution D on $\mathcal{V}$ such that $b^D_{\mathcal{V}} = b$ is not unique, but from Proposition 30 it follows easily that the associated SSD $m^D_{\mathcal{V}}$ is unique, thus I shall denote this with $m^b_{\mathcal{V}}$.

The belief in a proposition is the amount of the support we give to it, but not supporting a proposition does not mean impugning it: to impugn it we have to support its negation. Thus, given a belief b and a formula φ on $\mathcal{V}$, it is sensible to consider $1 - b(\neg\varphi)$ as the plausibility of φ ($1 - b(\neg\varphi)$ can also be considered as the lack of certainty of $\neg\varphi$: the plausibility of φ is thus the lack of certainty of $\neg\varphi$).

Definition 34 *Given a belief b on $\mathcal{V}$, the associated plausibility is the function*

$$\begin{array}{rrcl} p^b : & \mathcal{L}_{\mathcal{V}} & \longrightarrow & [0,1] \\ & \varphi & \longmapsto & 1 - b(\neg\varphi) \,. \end{array}$$

Given a support distribution D on $\mathcal{U}$, the associated plausibility on $\mathcal{V}$ is $p^D_{\mathcal{V}} = p^{b^D_{\mathcal{V}}}$.

Proposition 35 is the analogue of Proposition 30 for plausibility instead of belief, thus I shall omit the proof.

Proposition 35 *Given a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$ and a formula ψ on $\mathcal{V}$,*

$$p^D_{\mathcal{V}}(\psi) = \sum_{s \cap s_{\mathcal{U}\cup\mathcal{V}}(\psi) \neq \emptyset} m^D_{\mathcal{U}\cup\mathcal{V}}(s) = \begin{cases} \sum\limits_{\varphi_i \not\leftrightarrow \neg\psi} \alpha_i + \left(1 - \sum\limits_{i=1}^{n} \alpha_i\right) & \text{if } \neg\psi \notin \mathcal{T}_{\mathcal{V}}, \\ 0 & \text{if } \neg\psi \in \mathcal{T}_{\mathcal{V}}. \end{cases}$$

Theorem 36 states the desirable property $p^b \geq b$ (to believe something more than what we consider it to be plausible would not be a form of correct reasoning) and gives a new interpretation for the concept of plausibility: the plausibility of a proposition is the maximum grade with which we can believe it without having to reduce our belief in other propositions.

Theorem 36 *Given a belief b and a formula φ on $\mathcal{V}$,*

- $p^b(\varphi) \geq b(\varphi)$;
- $(\exists b' \in B_{\mathcal{V}} \;\; (b'(\varphi) = x \text{ and } b' \geq b)) \;\Leftrightarrow\; x \in \left[b(\varphi), p^b(\varphi)\right]$.

Proof. Exploiting as usual Propositions 8 and 24:

- $1 = b(\top) = b((\varphi \vee \neg\varphi)) \geq b(\varphi) + b(\neg\varphi) - b((\varphi \wedge \neg\varphi)) =$
  $= b(\varphi) + b(\neg\varphi) - b(\bot) = b(\varphi) + b(\neg\varphi)$,
  therefore $p^b(\varphi) = 1 - b(\neg\varphi) \geq b(\varphi)$.

- If $x \notin \left[b(\varphi), p^b(\varphi)\right]$, then there are two possibilities:
  if $x < b(\varphi)$, then b' should satisfy the impossible condition $x = b'(\varphi) \geq b(\varphi)$, thus b' does not exist;
  if $x > p^b(\varphi)$, then b' should satisfy the impossible condition
  $x = b'(\varphi) = 1 - p^{b'}(\neg\varphi) \leq 1 - b'(\neg\varphi) \leq 1 - b(\neg\varphi) = p^b(\varphi)$,
  thus b' does not exist.
  It remains to show that for an $x \in \left[b(\varphi), p^b(\varphi)\right]$ such a b' exists:
  if $b(\varphi) = p^b(\varphi)$, then $b' = b$ satisfies the conditions;
  otherwise let $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ be the support distribution on $\mathcal{V}$ defined in the proof of Theorem 33 ($b_{\mathcal{V}}^D = b$ and $\sum_{i=1}^{n} \alpha_i = 1$) and define
  $\delta = \frac{x - b(\varphi)}{p^b(\varphi) - b(\varphi)}$, $I = \{i \mid \varphi_i \not\rightsquigarrow \varphi \text{ and } \varphi_i \not\rightsquigarrow \neg\varphi\}$ and
  $D' = \{\langle\varphi_i, \alpha_i\rangle \mid i \in \{1, \ldots, n\} \setminus I\} \cup \bigcup_{i \in I} \{\langle\varphi_i, (1-\delta)\alpha_i\rangle, \langle(\varphi_i \wedge \varphi), \delta\alpha_i\rangle\}$.
  D' is a support distribution on $\mathcal{V}$ and $b' = b_{\mathcal{V}}^D$ satisfies the conditions:
  using Propositions 30 and 35, $b(\varphi) \neq p^b(\varphi) \Rightarrow (\varphi \notin \mathcal{T}_{\mathcal{V}} \text{ and } \neg\varphi \notin \mathcal{T}_{\mathcal{V}})$,
  therefore $b'(\varphi) = \sum_{\varphi_i \rightsquigarrow \varphi} \alpha_i + \sum_{i \in I} \delta\alpha_i = b(\varphi) + \delta\left(p^b(\varphi) - b(\varphi)\right) = x$;
  for a $\psi \in \mathcal{L}_{\mathcal{V}}$, since $\varphi_i \rightsquigarrow \psi \Rightarrow (\varphi_i \wedge \varphi) \rightsquigarrow \psi$, it is easy to see that $b'(\psi) \geq b(\psi)$. ■

Theorem 37 *Given a support distribution D on $\mathcal{U}$, the associated plausibility on $\mathcal{V}$ has the following properties:*

- *given two formulas φ and ψ on $\mathcal{V}$, $\varphi \sim \psi \Rightarrow p_{\mathcal{V}}^D(\varphi) = p_{\mathcal{V}}^D(\psi)$;*
- $p_{\mathcal{V}}^D(\bot) = 0$;
- $p_{\mathcal{V}}^D(\top) = 1$;
- *given a finite nonempty set F of formulas on $\mathcal{V}$,*

$$p_{\mathcal{V}}^D\left(\bigwedge F\right) \leq \sum_{\emptyset \neq F' \subset F} (-1)^{|F'|+1} p_{\mathcal{V}}^D\left(\bigvee F'\right).$$

Proof. Using the definition of plausibility, Propositions 8 and 24 and Theorem 32, the first three properties follow directly;
whereas for the last one we have also to use Lemma 31

and to consider that $\neg \bigwedge F \sim \bigvee \overline{F}$ for $\overline{F} = \{\neg\varphi \mid \varphi \in F\}$:

$$\begin{aligned} p_{\mathcal{V}}^{D}\left(\bigwedge F\right) &= 1 - b_{\mathcal{V}}^{D}\left(\bigvee \overline{F}\right) \leq 1 - \sum_{\emptyset \neq \overline{F'} \subset \overline{F}} (-1)^{\left|\overline{F'}\right|+1} b_{\mathcal{V}}^{D}\left(\bigwedge \overline{F'}\right) = \\ &= 1 - \sum_{\emptyset \neq F' \subset F} (-1)^{|F'|+1} \left(1 - p_{\mathcal{V}}^{D}\left(\bigvee F'\right)\right) = \\ &= \sum_{F' \subset F} (-1)^{|F'|} + \sum_{\emptyset \neq F' \subset F} (-1)^{|F'|+1} p_{\mathcal{V}}^{D}\left(\bigvee F'\right) = \sum_{\emptyset \neq F' \subset F} (-1)^{|F'|+1} p_{\mathcal{V}}^{D}\left(\bigvee F'\right). \blacksquare \end{aligned}$$

Theorem 37 is the analogue of Theorem 32 for plausibility instead of belief: as done before for beliefs, we can define $P_{\mathcal{V}}$ as the set of the functions $p \in [0,1]^{\mathcal{L}_{\mathcal{V}}}$ which satisfy the four properties of Theorem 37. Theorem 38, the analogue of Theorem 33, states that these properties are sufficient to characterize an acceptable plausibility, in the sense that for every function $p \in P_{\mathcal{V}}$, there is a support distribution on $\mathcal{V}$ whose associated plausibility is p.

Theorem 38 *Given a function $p \in P_{\mathcal{V}}$, there is a support distribution D on $\mathcal{V}$ such that $p_{\mathcal{V}}^{D} = p$.*

Proof. With the same arguments of the proof of Theorem 37, it is easy to show that the function $b \in [0,1]^{\mathcal{L}_{\mathcal{V}}}$ such that $b(\varphi) = 1 - p(\neg\varphi)$ is a belief on $\mathcal{V}$ (and of course $p = p^{b}$). Therefore from Theorem 33 we obtain that there is a support distribution D on $\mathcal{V}$ such that $b_{\mathcal{V}}^{D} = b$; i.e., such that $p = p^{b_{\mathcal{V}}^{D}} = p_{\mathcal{V}}^{D}$. $\blacksquare$

Given a belief (or plausibility) on $\mathcal{U}$, since the associated SSD on $\mathcal{U}$ is unique, the inductive, strong and deductive conclusions on another set of propositional symbols $\mathcal{V}$ are completely defined.

Definition 39 *Given two sets of propositional symbols $\mathcal{U}$ and $\mathcal{V}$, a belief b on $\mathcal{U}$ and a grade α, define $\mathcal{C}_{\mathcal{V}}^{\alpha}(b) = \mathcal{C}_{\mathcal{V}}^{\alpha}(D)$, $\mathcal{S}_{\mathcal{V}}^{\alpha}(b) = \mathcal{S}_{\mathcal{V}}^{\alpha}(D)$ and $\mathcal{D}_{\mathcal{V}}(b) = \mathcal{D}_{\mathcal{V}}(D)$, where D is a support distribution on $\mathcal{U}$ such that $b_{\mathcal{U}}^{D} = b$.*

From the definitions of associated belief and associated plausibility it follows directly that if $\mathcal{V} \subset \mathcal{U}$, then

$$\begin{aligned} \mathcal{C}_{\mathcal{V}}^{\alpha}(b) &= \{\psi \in \mathcal{L}_{\mathcal{V}} \mid b(\psi) \geq \alpha\}, \\ \mathcal{S}_{\mathcal{V}}^{\alpha}(b) &= \left\{\psi \in \mathcal{L}_{\mathcal{V}} \mid b(\psi) \geq \alpha,\ p^{b}(\psi) = 1\right\}, \\ \mathcal{D}_{\mathcal{V}}(b) &= \{\psi \in \mathcal{L}_{\mathcal{V}} \mid b(\psi) = 1\}. \end{aligned}$$

1.8 The Dempster-Shafer theory of evidence

The beliefs and the plausibilities introduced in Section 1.7 are the object of the so-called “Dempster-Shafer theory of evidence”, also known as “belief functions theory”. To be more precise, given a belief b on $\mathcal{V}$, the corresponding belief function is the function $bel^b \in [0,1]^{\mathcal{P}(V_\mathcal{V})}$ such that $b = bel^b \circ s_\mathcal{V}$ (from Theorem 32 and Proposition 24 it follows that bel^b is well-defined); therefore, in analogy with the definition of SSD, we could call bel^b the associated semantical belief on $\mathcal{V}$. The same relation occurs between a plausibility and the corresponding plausibility function.

The theory of belief functions was originated in the second half of the sixties by Dempster's study of upper and lower probabilities in the context of statistical research (see in particular [4], where Dempster's rule of combination is introduced, and [5]). But the real starting point of the theory can be considered Shafer's 1976 monograph [26], which (as the author says in the preface) "offers a reinterpretation of Dempster's work, a reinterpretation that identifies his "lower probabilities" as epistemic probabilities or degrees of belief, takes the rule for combining such degrees of belief as fundamental, and abandons the idea that they arise as lower bounds over classes of Bayesian probabilities".

The Dempster-Shafer theory of evidence involves combining independent items of evidence by representing each individually by a belief function and using Dempster's rule to combine these. It is a generalization of the Bayesian theory of probability in the sense that (at least at a purely formal level) epistemic probabilities are a special case of belief functions and probabilistic conditioning is a special case of Dempster's rule (and it is also possible to formulate generalizations of Bayes' theorem for belief functions). The big advantage of belief functions in comparison with probabilities is the ease in the representation of ignorance (since not supporting a proposition does not mean impugning it: $b(\varphi) + b(\neg\varphi)$ can be smaller than 1). Thus belief functions can be seen as a practical and powerful tool to represent human reasoning (see for instance [30]): this can explain the broad interest in the Dempster-Shafer theory manifested by the artificial intelligence community since the second half of the eighties. This popularity has also led to intensive debates between the supporters of the belief function theory and of the Bayesian theory, respectively (see for example the two special issues of the International Journal of Approximate Reasoning of September/November 1990 and of May 1992).

These discussions could be protracted forever, since the two opposed theories describe epistemic states and therefore are practically not comparable (a comparison, however very questionable, could be given by the respective effectiveness in the field of artificial intelligence). Some (subjectivist) Bayesians claim that belief functions lack a justification such as the one of epistemic probability based on betting coherence (see for instance [3]), without considering that this is highly debatable (and I think that an analogous justification, at least as sensible as the probabilistic one, could be found for belief functions). In fact the only support for the idea that degrees of belief have always a probabilistic structure comes from the case in which the object of our interest are the results of a repeatable situation: here (ontological) probabilities have a clear frequentist interpretation, and it is reasonable (in particular in games of chance) to accept these (ontological) probabilities as our degrees of belief. But this does not mean that not accepting them is unreasonable, or that for unrepeatable situations our beliefs have to be probabilistically structured; and anyway, since epistemic probabilities are a special kind of belief functions, the particular case of repeatable situations does not contradict at all the Dempster-Shafer theory. As games of chance are a classical example for Bayesian probability, so the problems of combination of testimonies are a classical topic for belief functions (because of the possibility of representing the total ignorance): in 1689 already, George Hooper published "A calculation of the credibility of human testimony", where he formulated special cases of Dempster's rule, as did also Jakob Bernoulli in his "Ars Conjectandi" (1713, posthumously) and Johann Heinrich Lambert in his "Neues Organon" (1764); for a historical perspective and a comparison between the belief function and the Bayesian approaches to this kind of problems see [27] and [28].

In the introduction (Section 1.1) I have said that default logic can be considered as a special case of the present formalization of inductive inference: from a formal standpoint this statement does not make much sense, for the simple reason that the meaning of "default logic" is not unique. In fact there are many different theories which are denoted by this name: obviously they are all related, and form an important part of the even broader set of theories known under the name of "nonmonotonic logics". Default logics have been developed to represent the aspect of human reasoning that is based on an "assumption of normality"[10] (called "default reasoning"): if the situation in which we are interested seems to be "normal", then we draw some conclusions; of course, if then we notice that the situation is a particular one, we should withdraw the conclusions based on the "assumption of normality", thus this kind of reasoning is nonmonotonic. The representation of default reasoning is very problematic: each one of the different formalizations of this reasoning resolves some problems but still presents some others; depending on which problems are considered more important, one formalization of default logic is preferred to another. What I meant when I said that default logic can be considered as a special case of the present formalization was that the aim of default logic (the representation of default reasoning) can be reached using the belief functions; moreover, using the belief functions, we can resolve very easily some of the biggest difficulties of default logic. But for what concerns the particular problems of the representation of default reasoning, a comparison of the results of the use of belief functions with the results of default logic is not so simple, since for every problem it is not clear at all what should be the result (i.e., which result would better describe human reasoning). At first I intended to write a short section about this topic, but later I realized that such a section could not be at the same time sensible and short. Thus I have preferred to simply refer the reader to the literature of nonmonotonic logic and belief functions theory: for a complete description of the different formalizations of default logic see [20] (and for an introduction see the already cited [2]); whereas for an introduction to the modelling of default reasoning with belief functions see [35], and for a complete theory see [1].

[10] "Normality" must be interpreted as a synonym of "typicality" or "regularity", thus "normality" has nothing to do with the normal distribution considered in stochasitics.

I shall now consider some particular kinds of beliefs (respectively belief functions): certainties, simple supports, epistemic probabilities, and consonant beliefs. These are equivalent to the mathematical structures studied in other theories of reasoning, but for the moment I shall consider only the static aspect: to be allowed to consider these theories as special cases of the one presented, we have to compare both the static and the dynamic aspects.

1.9 Particular kinds of beliefs

1.9.1 Certainties

Certainties are the kind of beliefs considered in classical propositional logic; that is, they are the beliefs associated with the classical support distributions.

Definition 40 *Given a set of propositional symbols $\mathcal{U}$, a belief b on $\mathcal{U}$ is called certainty if there is a classical support distribution D on $\mathcal{U}$ such that $b = b_{\mathcal{U}}^{D}$. In other words, $b \in B_{\mathcal{U}}$ is a certainty if there is a formula φ on $\mathcal{U}$ such that $b = b_{\mathcal{U}}^{\langle\varphi,1\rangle}$ (where $b_{\mathcal{U}}^{\langle\varphi,1\rangle}$ is an abbreviation of $b_{\mathcal{U}}^{\{\langle\varphi,1\rangle\}}$): b is called the certainty of φ; if $\varphi \in \mathcal{T}_{\mathcal{U}}$, then $b = b_{\mathcal{U}}^{\emptyset}$: b is called the vacuous belief on $\mathcal{U}$.*

Let $b \in B_{\mathcal{U}}$ be the certainty of $\varphi \in \mathcal{L}_{\mathcal{U}}$; then for every positive grade α

$$\mathcal{C}_{\mathcal{V}}^{\alpha}(b) = \mathcal{S}_{\mathcal{V}}^{\alpha}(b) = \mathcal{D}_{\mathcal{V}}(b) = \{\psi \in \mathcal{L}_{\mathcal{V}} \mid \varphi \rightsquigarrow \psi\}.$$

That is, there are no non-trivial inductive conclusions which are not deductive (in fact we are in the field of classical propositional logic, where only deductive conclusions are allowed).

Certainties are very particular beliefs: since they are the beliefs considered in classical propositional logic, they are related to deductive reasoning, and, as it is stated in the introduction (Section 1.1), deductive reasoning is monotonic: if we have a certainty, nothing we can add to our knowledge will change it (apart from the particular case in which even the monotonicity of classical propositional logic is debatable).

1.9.2 Simple supports

Simple supports are the simplest non-classical beliefs: they are associated with support distributions $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ such that there is at most one $\alpha_i \neq 0$; i.e., simple supports assign support to at most one formula (and the rest to $\top$). Of course certainties are simple supports.

Definition 41 *Given a set of propositional symbols $\mathcal{U}$, a belief b on $\mathcal{U}$ is called simple support if there is a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$ such that $b = b_{\mathcal{U}}^{D}$ and $|\{i \in \{1, \ldots, n\} \mid \alpha_i \neq 0\}| \leq 1$. In other words, $b \in B_{\mathcal{U}}$ is a simple support if there are a formula φ on $\mathcal{U}$ and a grade α such that $b = b_{\mathcal{U}}^{\langle\varphi,\alpha\rangle}$ (where $b_{\mathcal{U}}^{\langle\varphi,\alpha\rangle}$ is an abbreviation of $b_{\mathcal{U}}^{\{\langle\varphi,\alpha\rangle\}}$): b is called the simple support α of φ.*

Let $b \in B_{\mathcal{U}}$ be the simple support α of $\varphi \in \mathcal{L}_{\mathcal{U}}$; then for every positive grade $\beta \leq \alpha$

$$\mathcal{C}_{\mathcal{V}}^{\beta}(b) = \mathcal{S}_{\mathcal{V}}^{\beta}(b) = \{\psi \in \mathcal{L}_{\mathcal{V}} \mid \varphi \rightsquigarrow \psi\},$$

whereas for every grade $\beta > \alpha$

$$\mathcal{C}_{\mathcal{V}}^{\beta}(b) = \mathcal{S}_{\mathcal{V}}^{\beta}(b) = \mathcal{D}_{\mathcal{V}}(b) = \mathcal{T}_{\mathcal{V}}(b).$$

That is, there can be non-trivial inductive conclusions which are not deductive, but every conclusion is strong.

Simple supports are the kind of beliefs considered by the studies on combination of testimonies: if we believe that a witness is reliable at degree α, then we assign this support to the proposition stated by him/her, whereas the rest $1 - \alpha$ is not assigned to anything, because an unreliable witness gives no information. This process of construction of a belief (the one about the proposition) from another one (the belief about the reliability of the witness) will be considered precisely and more generally later, and the problems of combination of testimony will also return (since they are the clearest and oldest examples of combinations of beliefs).

1.9.3 Epistemic probabilities

Epistemic probabilities are beliefs such that the corresponding belief functions are probability distributions (Shafer in [26] called these "Bayesian belief functions"). Given a belief b on $\mathcal{U}$, from Proposition 30 it follows that $bel^{b}(s) = \sum_{s' \subset s} m_{\mathcal{U}}^{b}(s')$; thus bel^{b} is a probability distribution if and only if $m_{\mathcal{U}}^{b}(s) = 0$ for every s with more than one element.

Definition 42 *Given a set of propositional symbols $\mathcal{U}$, a belief b on $\mathcal{U}$ is called epistemic probability if bel^{b} is a probability distribution on $(V_{\mathcal{U}}, \mathcal{P}(V_{\mathcal{U}}))$. In other words, $b \in B_{\mathcal{U}}$ is an epistemic probability if, for every $s \subset V_{\mathcal{U}}$, $|s| \neq 1 \Rightarrow m_{\mathcal{U}}^{b}(s) = 0$.*

Epistemic probabilities have a lot of particular properties; some of them completely define this class of beliefs: Proposition 43 shows two possible characterizations (the proof is very easy considering Propositions 30 and 35).

Proposition 43 *Given a belief b on $\mathcal{U}$, the following three assertions are equivalent:*

- *b is an epistemic probability;*
- *$p^b = b$ (the belief and the associated plausibility are the same);*
- *for every formula φ on $\mathcal{U}$, $b(\varphi) + b(\neg\varphi) = 1$ (not supporting a proposition means impugning it).*

Let $b \in B_{\mathcal{U}}$ be an epistemic probability; then for every grade α

$$\mathcal{S}^{\alpha}_{\mathcal{V}}(b) = \mathcal{D}_{\mathcal{V}}(b).$$

That is, there are no strong conclusions which are not deductive (in Bayesian probability theory, strong means certain).

Epistemic probabilities are a kind of beliefs with which the standard methods of the Dempster-Shafer theory (in particular Dempster's rule of combination) have some problems. I shall try to reduce these problems later, but for the moment I would like to reconsider the importance of epistemic probabilities as beliefs: their use is based on the so called "principle of direct inference" (or "Hacking's principle"), which states the following: "If a person is certain that the ontological probability of the event A is p, then, if he/she is rational, his degree of belief of A's holding, should also be p". But what does "the ontological probability of the event A is p" mean? Beyond the philosophical subtleties, the only aspect of an ontological probability that everyone accepts is related to the long run frequencies (apart from the cases in which the classical definition of probability is applicable). Thus if A is a particular result of a repeatable situation B, then (following Hempel in [14]) "the ontological probability of the event A is p" means that in a long series of repetitions of B it is practically certain that the relative frequency of the result A will be approximately equal to p. The meaning of the words "is practically certain that" is clear: they shows that the logical connection between assertions about ontological probability and the associated assertions about empirical frequencies is inductive. Therefore, the assumption of the existence of an unique ontological probability is highly debatable, and even accepting this assumption, how can a person be certain that this probability is p? He/she can in the best case be certain that it is very close to p: how close depends on the quality of the information leading to this certainty. So I think that the principle of direct inference does not make so much sense: instead of assuming p as the degree of belief of A's holding, it is more rational to use a degree a little smaller than p; since this has to be done also for the event "not A", the belief and the plausibility will be the limits of an interval containing p, and, as it follows from the above reasoning, an interval is surely a better description of the belief than a single value (I shall consider more precisely this situation later, when I shall introduce the notion of discounting). Of course there are extreme situations, for instance the toss of a coin: in this case assuming $\frac{1}{2}$ and $\frac{1}{2}$ as the beliefs on head and tail, respectively, could be rational; but this is the standard example of randomness, there is a huge statistical knowledge about it, and above all there is something more: the symmetry (i.e., the classical definition of probability is applicable). I cannot discuss here the importance of symmetry in human reasoning, but clearly it is fundamental; just think about electronic games of chance, such as slot machines: they do not need symmetry to create randomness, but all of them simulate it, because otherwise no one would play.

1.9.4 Consonant beliefs

Consonant beliefs are equivalent to the measures of necessity studied in possibility theory (see [7]): they are associated to support distributions $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ such that $\varphi_i \rightsquigarrow \varphi_{i+1}$ (for every $i \in \{1, \ldots, n-1\}$). Of course simple supports are consonant beliefs, and also the belief associated with the support distribution of Example 5 is consonant.

Definition 44 *Given a set of propositional symbols $\mathcal{U}$, a belief b on $\mathcal{U}$ is called consonant if there is a support distribution $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ on $\mathcal{U}$ such that $b = b_{\mathcal{U}}^D$ and, for every $i \in \{1, \ldots, n-1\}$, $\varphi_i \rightsquigarrow \varphi_{i+1}$.*

As we shall see better later, consonant beliefs are in some senses opposite to epistemic probabilities, and like these they have a lot of particular properties, some of which completely define the consonance: Proposition 45 shows two possible characterizations of consonant beliefs (for the proof see Shafer's monograph [26]).

Proposition 45 *Given a belief b on $\mathcal{U}$, the following three assertions are equivalent:*

- *b is consonant;*
- *for every two formulas φ, ψ on $\mathcal{U}$, $b((\varphi \wedge \psi)) = \min(b(\varphi), b(\psi))$;*
- *for every two formulas φ, ψ on $\mathcal{U}$, $p^b((\varphi \vee \psi)) = \max\left(p^b(\varphi), p^b(\psi)\right)$.*

In possibility theory a consonant belief b is called necessity measure, whereas the associated plausibility p^b is the corresponding possibility measure. From Proposition 45 follows in particular that, for every formula φ on $\mathcal{U}$, since $(\varphi \wedge \neg\varphi) \sim \perp$, $\min(b(\varphi), b(\neg\varphi)) = 0$: two opposite propositions cannot both be the slightest bit necessary at the same time. Therefore a relation much stronger than $b \geq p^b$ is valid for consonant beliefs:

$$\begin{aligned} b(\varphi) > 0 \quad &\Rightarrow \quad p^b(\varphi) = 1; \\ p^b(\varphi) < 1 \quad &\Rightarrow \quad b(\varphi) = 0. \end{aligned}$$

This property means that a proposition is always completely possible before being in the least necessary.

Let $b \in B_{\mathcal{U}}$ be a consonant belief; then for every positive grade α

$$\mathcal{C}_{\mathcal{V}}^{\alpha}(b) = \mathcal{S}_{\mathcal{V}}^{\alpha}(b).$$

That is, every non-trivial conclusion is strong (in possibility theory, strong does not mean anything): in this sense consonant beliefs are opposite to epistemic probabilities.

The results of Proposition 45 show that consonant beliefs have a simple structure: this can be easily described by a fuzzy set (in fact normalized fuzzy sets and possibility measures are equivalent). Computing with fuzzy sets is relatively simple, thus if possibility measures and consonant beliefs would behave in the same way at the dynamic level, then it could be useful to approximate any belief function by a consonant one (see [9]); but unfortunately this is not the case: the two theories differ in the dynamic aspect.

2 The nonmonotonic evolution of a belief

I shall now consider the dynamic aspect of the present theory (and therefore of the Dempster-Shafer theory of evidence): the evolution of a belief due to the acquisition of new information. We shall see that this evolution is nonmonotonic: new knowledge can compel us to withdraw some of our former conclusions; but first I shall present the simple situation in which we merely extend or restrict the topic of interest.

2.1 Extensions and restrictions

Consider two sets of propositional symbols $\mathcal{U}$ and $\mathcal{V}$, such that $\mathcal{U} \subset \mathcal{V}$ (that is, $\mathcal{V}$ represent an extension of the topic represented by $\mathcal{U}$). A support distribution on $\mathcal{U}$ is also one on $\mathcal{V}$, but this is not valid for beliefs (since a belief on $\mathcal{U}$ is a function from $\mathcal{L}_{\mathcal{U}}$ to $[0, 1]$), thus the following definitions can be very useful.

Definition 46 *Given two sets of propositional symbols $\mathcal{U} \subset \mathcal{V}$, and two beliefs $b_{\mathcal{U}}$ and $b_{\mathcal{V}}$ on $\mathcal{U}$ and $\mathcal{V}$, respectively, $b_{\mathcal{V}}$ is an extension of $b_{\mathcal{U}}$ to $\mathcal{V}$ if $b_{\mathcal{U}} = b_{\mathcal{V}}{\downarrow_{\mathcal{U}}}$ ($b_{\mathcal{V}}{\downarrow_{\mathcal{U}}}$ is called the restriction of $b_{\mathcal{V}}$ to $\mathcal{U}$).*

Of course, the restriction of a belief on $\mathcal{V}$ to $\mathcal{U}$ is a belief on $\mathcal{U}$; and the restriction is unique, whereas there are many possible extensions of a belief (unless $\mathcal{U} = \mathcal{V}$, obviously). It is easy to show that if $b_{\mathcal{V}}$ is an extension of $b_{\mathcal{U}}$, then, for every grade α,

$$\begin{aligned} \mathcal{W} \subset \mathcal{U} \quad &\Rightarrow \quad \mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{U}}) = \mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{V}}) , \\ \mathcal{W} \not\subset \mathcal{U} \quad &\Rightarrow \quad \mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{U}}) \subset \mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{V}}) ; \end{aligned}$$

and in general if $\mathcal{W} \not\subset \mathcal{U}$, then $\mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{U}}) \neq \mathcal{C}^{\alpha}_{\mathcal{W}}(b_{\mathcal{V}})$. In other words, an extension of a belief carries the same information about the common topics, but can say more about other topics. Therefore it is useful to define the particular extension which does not carry any additional information: this is called the vacuous extension of a belief.

Definition 47 *Given two sets of propositional symbols $\mathcal{U} \subset \mathcal{V}$ and a belief b on $\mathcal{U}$, the vacuous extension of b to $\mathcal{V}$ is the belief $b{\uparrow_{\mathcal{V}}} = b^D_{\mathcal{V}}$, where D is a support distribution on $\mathcal{U}$ such that $b^D_{\mathcal{U}} = b$.*

For any set of propositional symbols $\mathcal{W}$, from the definition of $\mathcal{C}^{\alpha}_{\mathcal{W}}(b)$ follows directly that $\mathcal{C}^{\alpha}_{\mathcal{W}}(b{\uparrow_{\mathcal{V}}}) = \mathcal{C}^{\alpha}_{\mathcal{W}}(b)$; and Theorem 33 assures that the vacuous extension of a belief is always defined. It is easy to see that the vacuous extensions of certainties, simple supports or consonant beliefs are still certainties, simple supports or consonant beliefs, respectively, whereas the vacuous extension of an epistemic probability is not any more one of these (unless $\mathcal{U} = \mathcal{V}$, obviously); in fact among the four definitions of Section 1.9, the one of epistemic probability is the only one not based on support distributions. Therefore also in this sense consonant beliefs are opposite to epistemic probabilities (remember that certainties and simple supports are special cases of consonant beliefs).

2.2 Combinations of beliefs

The core of the Dempster-Shafer theory of evidence is Dempster's rule of combination for belief functions; these are considered as based on bodies of evidence, and the problem of combination can be formulated as follows: given two belief functions based on two bodies of evidence, how must we combine them to obtain the one based on the union of the two bodies? Dempster's rule describes how to combine two belief functions based on two independent bodies of evidence, where the meaning of "independent" can be explained through the probabilistic one if Dempster-Shafer theory is regarded as being founded on Bayesian probability theory (as in the original work of Dempster, [4], or in the late studies of Shafer: see for instance [29]), otherwise it can be explained by considering extensions of Dempster's rule (see for example [8]): "independence" becomes the name of a particular situation (like in probability theory). I shall follow the latter way: generalize Dempster's rule of combination and reconsider its importance by regarding it as the special situation of independence. As I have said in the introduction (Section 1.1), the theory presented stays between the two incompatible particular cases of classical propositional logic and Bayesian probability: therefore I shall consider the combination of beliefs in these two theories.

In classical propositional logic the combination of beliefs is very simple: if from a source of information (I prefer to speak about sources of information instead of bodies of evidence, but this does not change anything at the practical level) I get the certainty of φ and from another the certainty of ψ, then from both together I get the certainty of $(\varphi \wedge \psi)$.[11] In the present notation, a belief is described with a support distribution $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$: for purely formal reasons, during the study of the problem of combination, it is useful to consider only complete support distributions, where "complete" means that $\sum_{i=1}^{n} \alpha_i = 1$. Of course, every support distribution can be completed, in the sense that, instead of using $\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$, we can consider the equivalent complete support distribution $\left\{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle, \left\langle\top, 1 - \sum_{i=1}^{n} \alpha_i\right\rangle\right\}$: thus the requirement of the completeness is

[11] As I have said in a footnote on page 4, in the standard notation of classical propositional logic a belief is expressed by a set of formulas (the meaning is the certainty of their conjunction), and the combination of two beliefs is expressed by considering the union of the corresponding two sets of formulas: this is in accordance with what I have written.

not a restriction.

Of course, in the present notation the classical combination can be described as follows: given two complete classical support distributions $\{\langle\varphi_i,\alpha_i\rangle \mid i\in\{1,\ldots,n\}\}$ and $\big\{\big\langle\psi_j,\beta_j\big\rangle \mid j\in\{1,\ldots,m\}\big\}$, their combination is the (classical) support distribution

$$\big\{\big\langle(\varphi_i\wedge\psi_j),\alpha_i\beta_j\big\rangle \mid i\in\{1,\ldots,n\}\,,j\in\{1,\ldots,m\}\big\}$$

(in fact, since $\alpha_i,\beta_j\in\{0,1\}$, $\alpha_i\beta_j\neq 0 \Leftrightarrow \alpha_i=\beta_j=1$). This could be inconsistent: in classical propositional logic this is not regarded as a problem, but if we would like to generalize this rule of combination directly by allowing any complete support distribution, since we do not accept inconsistency, we should (if possible) renormalize it, thus obtaining:

$$\overline{\big\{\big\langle(\varphi_i\wedge\psi_j),\alpha_i\beta_j\big\rangle \mid i\in\{1,\ldots,n\}\,,j\in\{1,\ldots,m\}\big\}}. \tag{1}$$

This is a syntactical definition of Dempster's rule of combination, enlightening on its simplicity and clarity. The problem of the possible inconsistency and the consequent renormalization are very important: they are the source of the nonmonotonicity; I shall consider this topic later. Although the given expression of the classical combination in the present notation is the most immediate one, this could also be formulated as follows: given two complete classical support distributions $\{\langle\varphi_i,\alpha_i\rangle \mid i\in\{1,\ldots,n\}\}$ and $\big\{\big\langle\psi_j,\beta_j\big\rangle \mid j\in\{1,\ldots,m\}\big\}$, their combination is the (classical) support distribution

$$\begin{gathered}\big\{\big\langle(\varphi_i\wedge\psi_j),\gamma_{ij}\big\rangle \mid i\in\{1,\ldots,n\}\,,j\in\{1,\ldots,m\}\big\}\,\text{, such that}\\ \sum_{j=1}^{m}\gamma_{ij}=\alpha_i \text{ for every } i\in\{1,\ldots,n\}\,\text{, and } \sum_{i=1}^{n}\gamma_{ij}=\beta_j \text{ for every } j\in\{1,\ldots,m\}\end{gathered}$$

(in fact it is easy to see that, among the γ_{ij}, exactly one has value 1, and all the others have value 0). As before, we can generalize this rule of combination by allowing any complete support distribution and renormalizing:

$$\begin{gathered}\overline{\big\{\big\langle(\varphi_i\wedge\psi_j),\gamma_{ij}\big\rangle \mid i\in\{1,\ldots,n\}\,,j\in\{1,\ldots,m\}\big\}}\text{, such that}\\ \sum_{j=1}^{m}\gamma_{ij}=\alpha_i \text{ for every } i\in\{1,\ldots,n\}\,\text{, and } \sum_{i=1}^{n}\gamma_{ij}=\beta_j \text{ for every } j\in\{1,\ldots,m\}\,;\end{gathered} \tag{2}$$

but, unlike the classical case, in general the values of the γ_{ij} must be chosen.

In Bayesian probability theory, only one special case of combination of beliefs is universally accepted: the conditioning of an epistemic probability on a certainty. If we express it in the present notation, this becomes obviously a special case of Dempster's rule (1), but also of the more general rule (2): in fact it is easy to see that, in the particular case of conditioning, the γ_{ij} are completely determined by the conditions α_i and β_j. As in classical propositional logic, in the Bayesian theory the only beliefs one can get from new pieces of information are certainties: if we want to model human reasoning, this is unsatisfactory, because hardly ever we can get a certainty from a piece of information (remember the particularity of a certainty, as the only kind of belief we shall never withdraw). Thus we should extend the classical and Bayesian rules of combination to a single rule accepting every kind of beliefs without distinction.

The combinations of beliefs in classical propositional logic and Bayesian probability theory are merely automatic operations: given two beliefs, we obtain directly their combination, without considering the meaning of the formulas involved. We could assume this as a property which should be satisfied by any rule of combination: with such an assumption, Dempster's is the only sensible extension of the classical and Bayesian rules; in fact Dubois and Prade in [8] have shown that, from the hypothesis $\gamma_{ij} = \alpha_i \star \beta_j$ with $\star$ a continuous operation, follows that $\star$ must be the product. But Dempster's rule has a big problem: to avoid absurdities, the notion of the independence of the sources of information, which is very hard to be defined, must be introduced. Alternatively, we could consider the automaticity as a feature of the special case of combining a belief with a certainty, and choose the values of the γ_{ij} by considering the meaning of the formulas involved: this allows a simple definition of independence. For this reason and for the sake of generality, I shall follow the latter way and therefore consider the general rule (2).

Consider a topic $\mathcal{U}$ and two sources of information, from which I get two beliefs b_1 and b_2 on $\mathcal{U}$. The assumption that the two beliefs are on the same set of propositional symbols, is not a restriction, since if the two beliefs b_1 and b_2 were on $\mathcal{U}_1$ and $\mathcal{U}_2$, respectively, then we could consider their vacuous extension on $\mathcal{U} = \mathcal{U}_1 \cup \mathcal{U}_2$. To avoid absurdities, it is very important to assume that the two beliefs are the most specific possible descriptions of what the sources of information told me about $\mathcal{U}$. The rule (2) is about support distributions, but since a belief can be described by many of them, it is more convenient to consider the description which is uniquely related to a belief: the associated SSD.

Definition 48 *Given two beliefs b_1 and b_2 on $\mathcal{U}$, a joint semantical support distribution (JSSD) with marginal beliefs b_1 and b_2 is a function*

$$\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2} : \mathcal{P}(V_{\mathcal{U}}) \times \mathcal{P}(V_{\mathcal{U}}) \longrightarrow [0,1], \text{ such that}$$

$$\sum_{s' \subset V_{\mathcal{U}}} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s, s') = m_{\mathcal{U}}^{b_1}(s) \text{ for every } s \subset V_{\mathcal{U}}, \text{ and}$$
$$\sum_{s \subset V_{\mathcal{U}}} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s, s') = m_{\mathcal{U}}^{b_2}(s') \text{ for every } s' \subset V_{\mathcal{U}}.$$

The conflict of b_1 and b_2 with respect to a JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ is

$$c(b_1 \otimes b_2) = \sum_{s \cap s' = \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s, s');$$

the JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ *is called acceptable if* $c\,(b_1 \otimes b_2) < 1$.
The combination of b_1 *and* b_2 *with respect to an acceptable JSSD* $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ *is the belief on* $\mathcal{U}$

$$\begin{array}{rcll} b_1 \otimes b_2 : & \mathcal{L}_{\mathcal{U}} & \longrightarrow & [0,1] \\ & \varphi & \longmapsto & \frac{1}{1-c(b_1 \otimes b_2)} \sum\limits_{\emptyset \neq s \subset s_{\mathcal{U}}(\varphi)} \sum\limits_{s' \cap s'' = s} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\,(s', s'')\,. \end{array}$$

Of course the choice of a JSSD with the given marginal beliefs reflects the choice of the γ_{ij} in the syntactical definition (2) of the rule, whereas the conflict is the degree of inconsistence of the obtained support distribution before the renormalization, which is possible if the JSSD is acceptable. It is easy to see that, for the SSD associated with $b_1 \otimes b_2$, the following is valid:

$$m_{\mathcal{U}}^{b_1 \otimes b_2}\,(s) = \begin{cases} \frac{1}{1-c(b_1 \otimes b_2)} \sum\limits_{s' \cap s'' = s} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\,(s', s'') & \text{if } s \neq \emptyset, \\ 0 & \text{if } s = \emptyset. \end{cases}$$

Dempster's rule of combination is a special case of the formulated one, because $\underline{m}_{\mathcal{U}}^{b_1 \oplus b_2}\,(s, s') = m_{\mathcal{U}}^{b_1}\,(s)\, m_{\mathcal{U}}^{b_2}\,(s')$ is obviously a JSSD with marginal beliefs b_1 and b_2; the respective conflict of b_1 and b_2 is denoted by $c\,(b_1 \oplus b_2)$, and the respective combination of b_1 and b_2 by $b_1 \oplus b_2$. Proposition 49 states that Dempster's rule is always applicable, when the general one is applicable; and they are inapplicable only in the extreme situation in which a proposition is considered certain from the first source of information and impossible from the latter one (I shall speak about this situation later); the proof is simple.

Proposition 49 *Given two beliefs* b_1 *and* b_2 *on* $\mathcal{U}$*, the following three assertions are equivalent:*

- $\underline{m}_{\mathcal{U}}^{b_1 \oplus b_2}$ *is not acceptable;*
- *no JSSD* $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ *is acceptable;*
- *there is a formula* φ *on* $\mathcal{U}$*, such that* $b_1\,(\varphi) = 1$ *and* $p^{b_2}\,(\varphi) = 0$.

Example 50 *As I have said in Section 1.8, the first example of a rule which is identifiable as a special case of Dempster's one appears in Hooper's studies on the problems of combination of testimonies. It is the "rule for concurrent testimony", which states that if a report is concurrently attested by* n *independent reporters, each with reliability* $\alpha \in [0,1]$*, then the credibility of the report is* $1 - (1-\alpha)^n$*. I think that in this case the meaning of "independent" is clear: it means simply that the reporters have nothing to do with each other. We can model this situation by considering a single propositional symbol* r *(which represents the report) and the beliefs* $b_1, \ldots, b_n$ *on* $\{r\}$ *that we get from each* *testimony; obviously these beliefs are equal: as it is stated in Section 1.9, every b_i is the simple support α on r. It is easy to calculate their combination: $b_1 \oplus \ldots \oplus b_n = b_{\{r\}}^{\langle r, 1-(1-\alpha)^n \rangle}$ (I have omitted the brackets, since Dempster's rule is associative: this is a consequence of the associativity of the product); therefore $(b_1 \oplus \ldots \oplus b_n)(r) = 1 - (1-\alpha)^n$, as it is stated by Hooper's rule for concurrent testimony.*

The given definition of combination raises a fundamental question: if I get two beliefs b_1 and b_2 on $\mathcal{U}$ from two sources of information, how should I choose the JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$? First of all, it is important to notice that, if one of the two beliefs is a certainty, then there is nothing to choose, since the JSSD is unique (this is the situation of the classical and Bayesian combinations): $b \otimes b_{\mathcal{U}}^{\langle \varphi, 1 \rangle} = b \oplus b_{\mathcal{U}}^{\langle \varphi, 1 \rangle}$, and $c\left(b \otimes b_{\mathcal{U}}^{\langle \varphi, 1 \rangle}\right) = b(\neg\varphi)$ (analogously if the certainty is the first marginal belief). In particular, if the certainty is the vacuous belief, then $c\left(b \otimes b_{\mathcal{U}}^{\emptyset}\right) = 0$ and $b \otimes b_{\mathcal{U}}^{\emptyset} = b$: the vacuous belief is a sort of neutral element of the combination. But in general the question raised is difficult, and to get an answer it is very useful to introduce the notion of conditional belief.

Definition 51 *Given two sets of propositional symbols $\mathcal{U}$ and $\mathcal{V}$, a conditional belief on $\mathcal{V}$ given $\mathcal{U}$ is a mapping*

$$\begin{array}{rccl} b \mid \mathcal{U}: & \mathcal{L}_{\mathcal{U}} & \longrightarrow & B_{\mathcal{V}} \\ & \varphi & \longmapsto & b \mid \varphi, \end{array}$$

such that for all formulas φ, ψ on $\mathcal{U}$: $\varphi \sim \psi \Rightarrow b \mid \varphi = b \mid \psi$.

Obviously, given a belief b on $\mathcal{U}$ and a conditional belief $b \mid \mathcal{U}$ on $\mathcal{U}$, the function $\underline{m}_{\mathcal{U}}^{b \otimes (b|\mathcal{U})}(s, s') = m_{\mathcal{U}}^{b}(s)\, m_{\mathcal{U}}^{b|\varphi_s}(s')$ (where, for every $s \subset V_{\mathcal{U}}$, φ_s is a formula on $\mathcal{U}$ such that $s_{\mathcal{U}}(\varphi_s) = s$) is a JSSD with marginal beliefs b and $b' = \sum_{s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b}(s)(b \mid \varphi_s)$ (a weighted average of belief is still a belief); the respective conflict of b and b' is denoted by $c(b \otimes (b \mid \mathcal{U}))$, and the respective combination of b and b' by $b \otimes (b \mid \mathcal{U})$. Conversely, it is easy to see that, given a JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$, there is a conditional belief $b \mid \mathcal{U}$ on $\mathcal{U}$, such that $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2} = \underline{m}_{\mathcal{U}}^{b_1 \otimes (b|\mathcal{U})}$.

Therefore, if I get two beliefs b_1 and b_2 on $\mathcal{U}$ from two sources of information, I can construct the JSSD by defining a conditional belief $b \mid \mathcal{U}$ on $\mathcal{U}$ such that $\sum_{s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1}(s)(b \mid \varphi_s) = b_2$; where, for every formula φ on $\mathcal{U}$, $b \mid \varphi$ must be interpreted as the belief that I would get from the second source of information if the first piece of information were the certainty of φ (in fact $b_{\mathcal{U}}^{\langle \varphi, 1 \rangle} \otimes (b \mid \mathcal{U}) = b_{\mathcal{U}}^{\langle \varphi, 1 \rangle} \otimes (b \mid \varphi)$). Of course, two sources of information are dependent if the belief that I would get from one of them depends on the piece of information yielded by the other; thus the two sources of information are independent if $b \mid \mathcal{U}$ is constant. With this definition of independence, we obtain the following: if I get two beliefs b_1 and b_2 on $\mathcal{U}$ from two independent sources of information, then $b_1 \otimes b_2 = b_1 \otimes (b \mid \mathcal{U}) = b_1 \oplus b_2$, since from $b \mid \mathcal{U}$ constant follows $b_2 = \sum_{s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1}(s)\,(b \mid \varphi_s) = b \mid \varphi$ for every formula φ on $\mathcal{U}$, and therefore $\underline{m}_{\mathcal{U}}^{b_1 \otimes (b|\mathcal{U})}(s, s') = m_{\mathcal{U}}^{b_1}(s)\, m_{\mathcal{U}}^{b_2}(s') = \underline{m}_{\mathcal{U}}^{b_1 \oplus b_2}(s, s')$. At the other extreme is the situation in which the sources of information are the same (from the assumption that the two beliefs are the most specific possible descriptions of what the sources of information told me about $\mathcal{U}$, it follows then that $b_1 = b_2$): in this case, for every formula φ on $\mathcal{U}$, $b \mid \varphi$ must be the certainty of φ (this is possible, since $b_2 = \sum_{s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1}(s)\,(b \mid \varphi_s) = \sum_{s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1}(s)\, b_{\mathcal{U}}^{\langle \varphi, 1 \rangle} = b_1$, which follows easily by reasoning on $m_{\mathcal{U}}^{b_1}$), therefore $b_1 \otimes b_2 = b_1 \otimes (b \mid \mathcal{U}) = b_1$, since $\underline{m}_{\mathcal{U}}^{b_1 \otimes (b|\mathcal{U})}(s, s') = m_{\mathcal{U}}^{b_1}(s)\, m_{\mathcal{U}}^{\langle \varphi_s, 1 \rangle}(s') = \begin{cases} m_{\mathcal{U}}^{b_1}(s) & \text{if } s = s', \\ 0 & \text{if } s \neq s', \end{cases}$ and so $m_{\mathcal{U}}^{b_1 \otimes (b|\mathcal{U})} = m_{\mathcal{U}}^{b_1}$.

Example 52 *Consider the following situation: I have heard the weather forecasts from two different radio stations, and both have said that it is likely that tomorrow it will rain in Zürich, but perhaps the first one has been more convincing. The beliefs that I have got from the two radio stations are the simple supports α and β, respectively, on r, where r represents the proposition "Tomorrow it will rain in Zürich" and $\alpha \geq \beta$; which is the belief (on $\{r\}$) that I have got from both radio stations together? Since almost certainly these have broadcast the weather prediction of the same meteorological station, the independence assumption would be absurd; therefore we have to choose a JSSD $\underline{m}_{\{r\}}^{b_{\{r\}}^{\langle r,\alpha \rangle} \otimes b_{\{r\}}^{\langle r,\beta \rangle}}$. This choice is equivalent with the one of a conditional belief $b \mid \{r\}$ on $\{r\}$ such that $\alpha\,(b \mid r) + (1-\alpha)\,(b \mid \top) = b_{\{r\}}^{\langle r,\beta \rangle}$; this equation is valid if and only if $b \mid r = b_{\{r\}}^{\langle r,\gamma \rangle}$ and $b \mid \top = b_{\{r\}}^{\left\langle r, \frac{\beta - \alpha\gamma}{1-\alpha} \right\rangle}$ for a γ such that* $\max\left(1 - \frac{1-\beta}{\alpha}, 0\right) \leq \gamma \leq \frac{\beta}{\alpha}$*. Thus γ is the belief that I think I would have got from the second radio station if the first one had said that tomorrow it will certainly rain in Zürich. If I were certain that the two stations have forecast the same weather prediction, then I would choose $\gamma = \frac{\beta}{\alpha}$ (which in this case would be the ratio of the convincingness of the two weather forecasters), and thus obtain the combined belief $b_{\{r\}}^{\langle r,\alpha \rangle} \otimes b_{\{r\}}^{\langle r,\beta \rangle} = b_{\{r\}}^{\langle r,\alpha \rangle}$ (the belief in r would be the maximum of the ones that I have got from the two sources of information). Otherwise, if I were less sure, I would choose a γ such that $\beta \leq \gamma \leq \frac{\beta}{\alpha}$ (where β would be the case of the independence), and thus obtain the combined belief $b_{\{r\}}^{\langle r,\alpha \rangle} \otimes b_{\{r\}}^{\langle r,\beta \rangle} = b_{\{r\}}^{\langle r,\beta + (1-\gamma)\alpha \rangle}$ (i.e., my belief in r would be in $[\alpha, \alpha + \beta - \alpha\beta]$). In this example the choice of a γ smaller than β is senseless, since it means that the belief in r that I have got from the second radio station would have been smaller if the first one had said that the rain is certain. However, formally the range of the possible choices of γ is $\left[\max\left(1 - \frac{1-\beta}{\alpha}, 0\right), \frac{\beta}{\alpha}\right]$: when α tends to 1, this range converges toward the singleton $\{\beta\}$; this can be read as follows: the more we approach a certainty, the less the dependence and the independence are important (and in the extreme case of a certainty, they are of no importance).*

In the Dempster-Shafer theory of evidence it is assumed that every kind of knowledge we get from a source of information can be described with a belief function. In the present formulation, it is also allowed to describe a piece of information with a conditional belief: this generalization permits to deal also with those situations in which the knowledge obtained from a source of information describes the dependence of a topic on a different one, about which we have already a belief; for example the dependence of an event on a previous one. Consider also two topics $\mathcal{U}$ and $\mathcal{V}$ such that $\mathcal{U}\cap\mathcal{V}=\emptyset$, and two sources of information such that from the first I get a belief b on $\mathcal{U}$ and from the latter a conditional belief $b \mid \mathcal{U}$ on $\mathcal{V}$: from both together I get the belief $(b{\uparrow_{\mathcal{U}\cup\mathcal{V}}}) \otimes (b' \mid \mathcal{U}\cup\mathcal{V})$, where $b' \mid \mathcal{U}\cup\mathcal{V}$ is the conditional belief on $\mathcal{U}\cup\mathcal{V}$ defined through $b' \mid \varphi = \begin{cases} (b \mid \varphi){\uparrow_{\mathcal{U}\cup\mathcal{V}}} & \text{if } \varphi \in \mathcal{L}_{\mathcal{U}}, \\ b^{\emptyset}_{\mathcal{U}\cup\mathcal{V}} & \text{if } \varphi \in (\mathcal{L}_{\mathcal{U}\cup\mathcal{V}} \setminus \mathcal{L}_{\mathcal{U}}) . \end{cases}$ Informally, we can consider $b' \mid \mathcal{U}\cup\mathcal{V}$ as a vacuous extension of the conditional belief $b \mid \mathcal{U}$; to simplify the notation, I shall write $b \otimes (b \mid \mathcal{U})$ instead of $(b{\uparrow_{\mathcal{U}\cup\mathcal{V}}}) \otimes (b' \mid \mathcal{U}\cup\mathcal{V})$, forgetting the vacuous extensions. From $\mathcal{U}\cap\mathcal{V}=\emptyset$ it follows easily that $c\,(b \otimes (b \mid \mathcal{U})) = 0$ and that $b\otimes(b \mid \mathcal{U})$ is an extension of b to $\mathcal{U}\cup\mathcal{V}$. It is important to notice that the analogy with the situation described above, in which the conditional belief was used to construct a JSSD, is only formal: $b \mid \varphi$ should not be interpreted as the belief that I would get from the second source of information if the first piece of information were the certainty of φ; $b \mid \varphi$ is a part of the knowledge obtained from the second source of information, it is the belief on $\mathcal{V}$ that the second source would give me if φ were true, and it is the belief on $\mathcal{V}$ that I would get from both sources together if the first piece of information were the certainty of φ. Therefore, if $b \mid \mathcal{U}$ is constant, it does not make sense to consider the two sources of information as independent: $b \mid \mathcal{U}$ constant rather signifies that the two topics $\mathcal{U}$ and $\mathcal{V}$ are independent in a probabilistically tasting sense. In fact this way of constructing an extension of a belief on $\mathcal{U}$ to $\mathcal{U}\cup\mathcal{V}$ through a conditional belief on $\mathcal{V}$ given $\mathcal{U}$ is a generalization of the process of defining a probability distribution on $V_{\mathcal{U}} \times V_{\mathcal{V}}$ through one on $V_{\mathcal{U}}$ and a stochastic kernel from $V_{\mathcal{U}}$ to $V_{\mathcal{V}}$ (since $V_{\mathcal{U}} \times V_{\mathcal{V}}$ and $V_{\mathcal{U}\cup\mathcal{V}}$ are identifiable in a natural way).

Example 53 *In Example 5 we considered a belief on $\{r, w\}$ (the topic of the weather situation of Lucerne); this belief can be seen as obtained from two sources of information: the first one tells me that it is very likely that it is raining in Lucerne, whereas the second one tells me that in this case it is likely that the wind is blowing. Thus from the first source of information I get the belief $b^{\langle r,0.9\rangle}_{\{r\}}$, whereas from the other one I get the conditional belief $b \mid \{r\}$ on $\{w\}$ such that $b \mid r = b^{\langle w,\frac{7}{9}\rangle}_{\{w\}}$ and, for every $\varphi \not\sim r$, $b \mid \varphi = b^{\emptyset}_{\{w\}}$. Therefore from both sources together I get $b \otimes (b \mid \{r\})$, which is the belief on $\{r, w\}$ associated with the support distribution described in Example 5.*

In the original work of Dempster, [4], upper and lower probabilities (which are mathematically equivalent to plausibility and belief functions, respectively) on a space S are constructed from a probability distribution on another space X and a multivalued mapping Γ from X to S (i.e., a mapping $\Gamma : X \longrightarrow \mathcal{P}(S)$). This process can be described (at least formally) by considering an epistemic probability b on $\mathcal{U}$ and a conditional belief $b \mid \mathcal{U}$ on $\mathcal{V}$ (where $\mathcal{U} \cap \mathcal{V} = \emptyset$) such that, for every formula φ on $\mathcal{U}$, $b \mid \varphi$ is a certainty. This can be seen by looking at the semantical counterpart of these objects: bel^b is a probability distribution on $V_{\mathcal{U}}$, whereas defining ψ_v as the formula on $\mathcal{V}$ such that $b \mid \varphi_{\{v\}}$ is the certainty of ψ_v (where as usual $\varphi_{\{v\}}$ is a formula on $\mathcal{U}$ such that $s_{\mathcal{U}}\left(\varphi_{\{v\}}\right) = \{v\}$),

$$\begin{array}{rrcl} \Gamma^{b|\mathcal{U}}: & V_{\mathcal{U}} & \longrightarrow & \mathcal{P}(V_{\mathcal{V}}) \\ & v & \longmapsto & s_{\mathcal{V}}(\psi_v) \end{array}$$

is a multivalued mapping from $V_{\mathcal{U}}$ to $V_{\mathcal{V}}$. It is easy to see that the lower probability on $V_{\mathcal{V}}$ induced by bel^b and $\Gamma^{b|\mathcal{U}}$ is the semantical counterpart of the belief $(b \otimes (b \mid \mathcal{U})) \downarrow_{\mathcal{V}}$, and that every belief on $\mathcal{V}$ can be constructed in this way.

Example 54 *Thanks to the syntactical version of a multivalued mapping (i.e., a conditional belief $b \mid \mathcal{U}$ on $\mathcal{V}$ such that $\mathcal{U} \cap \mathcal{V} = \emptyset$ and, for every formula φ on $\mathcal{U}$, $b \mid \varphi$ is a certainty), we can describe the construction of the belief that we get from a testimony better than what has been done in Section 1.9. If we have an epistemic probability $b_{\{r\}}^{\{\langle r,\alpha\rangle,\langle\neg r,1-\alpha\rangle\}}$ about the reliability of a witness (where r represents "the witness is reliable") and he/she states the proposition p, then the belief on $\{p\}$ that we shall obtain can be constructed by considering the conditional belief $b \mid \{r\}$ on $\{p\}$ such that $b \mid r = b_{\{p\}}^{\langle p,1\rangle}$ (if the witness were reliable, then p would be true) and $b \mid \neg r = b \mid \top = b_{\{p\}}^{\emptyset}$ (if the witness could be unreliable, then we could not say anything about p). Thus the belief on $\{p\}$ that we get from this testimony is $\left(b_{\{r\}}^{\{\langle r,\alpha\rangle,\langle\neg r,1-\alpha\rangle\}} \otimes (b \mid \{r\})\right) \downarrow_{\{p\}} = b_{\{p\}}^{\langle p,\alpha\rangle}$: the simple support α of p. It is important to notice that if our belief about the reliability of the witness were the simple support α of r instead of being the above stated epistemic probability (i.e., we give no support to the unreliability), the result would not change.*

Klawonn and Schwecke in [15] and Smets in [31] have shown that Dempster's rule of combination is the only one which satisfies a given set of properties; i.e., they have given an axiomatic justification of this rule: if we assume these properties as necessary, then we must combine beliefs with Dempster's rule. The sets of properties used by these authors are partially different: I shall consider the one proposed by Klawonn and Schwecke (since it is simpler and mathematically better formulated). Apart from two technical (and a little unjustified) assumptions, this set contains four properties: one states that the combination of a belief with a certainty must be the one obtained using Dempster's rule (this is also called Dempster's conditioning: I shall consider it later); whereas the other properties are the following (where $b_1, b_2, b_3 \in B_{\mathcal{U}}$):

- commutativity: $b_1 \oplus b_2 = b_2 \oplus b_1$;

- associativity: $(b_1 \oplus b_2) \oplus b_3 = b_1 \oplus (b_2 \oplus b_3)$;
- invariance under vacuous extension: $\mathcal{U} \subset \mathcal{V} \Rightarrow (b_1 \uparrow_{\mathcal{V}}) \oplus (b_2 \uparrow_{\mathcal{V}}) = (b_1 \oplus b_2) \uparrow_{\mathcal{V}}$.

The first property proposed by Klawonn and Schwecke (i.e., the conditioning property) is shared also by the generalization of Dempster's rule presented here; and the commutativity and the invariance under vacuous extension too, if the JSSD's are chosen in a sensible way (where "sensible" could be exactly defined, but I think that its meaning is intuitively clear). We could also impose conditions on the choice of the JSSD's so as to obtain the associativity (this can be seen by thinking of a three-dimensional JSSD), but I do not regard this property as necessary. Indeed, in my opinion, the associativity is unavoidable only if we have to combine a set of n beliefs which are not temporally ordered, and in this case it would be easier to generalize the rule of combination to n beliefs (by considering a n-dimensional JSSD); but if the beliefs are obtained and combined in a chronological order, I do not see why the resulting one should necessarily be the same if the chronological order were different. The property of Dempster's rule surely not shared by its generalization is what Smets calls "compositionality": for every set of propositional symbols $\mathcal{U}$, $\oplus \in {B_{\mathcal{U}}}^{B_{\mathcal{U}} \times B_{\mathcal{U}}}$ (in the discussion above, I called it "automaticity"). This property is necessary if we need a rule of combination which can be applied without having to reflect (for instance by a computer). But using Dempster's rule means assuming the independence of the sources of information; with the assumption of the independence, the conflict of the two beliefs can be very high, also if they are very similar or even the same. We shall see that the conflict is very unfavorable: it denotes how much a belief will be changed by combining it with the other. Therefore, if we do not know the relation between the two sources of information, it seems to me a lot more sensible to try to reduce the conflict of the respective beliefs: intuitively, this corresponds to the assumption of a "positive correlation" between the sources.

Given two beliefs b_1 and b_2 on $\mathcal{U}$, my idea is thus to choose a JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$, such that the respective conflict $c\,(b_1 \otimes b_2)$ is minimal or at least close to this minimum. For the minimal conflict of b_1 and b_2, the following inequality can be easily proved:

$$\min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\,(b_1 \otimes b_2) \geq \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_1\,(\varphi) - p^{b_2}\,(\varphi) \right);$$

I am convinced that the equality is also valid, but I have proved it only for some particular cases.

Conjecture 55 *Given two beliefs b_1 and b_2 on $\mathcal{U}$,*

$$\min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\,(b_1 \otimes b_2) = \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_1\,(\varphi) - p^{b_2}\,(\varphi) \right).$$

Thanks to the symmetry in the definition of JSSD, $\min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c(b_1 \otimes b_2) = \min_{\underline{m}_{\mathcal{U}}^{b_2 \otimes b_1}} c(b_2 \otimes b_1)$, and in fact also $\max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_1(\varphi) - p^{b_2}(\varphi)\right) = \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_2(\varphi) - p^{b_1}(\varphi)\right)$ is valid, because $b_1(\varphi) - p^{b_2}(\varphi) = \left(1 - p^{b_1}(\neg\varphi)\right) - (1 - b_2(\neg\varphi)) = b_2(\neg\varphi) - p^{b_1}(\neg\varphi)$. For instance, we have already noticed that, if one of the two beliefs is a certainty, then the JSSD is unique: $b \otimes b_{\mathcal{U}}^{\langle\psi,1\rangle} = b \oplus b_{\mathcal{U}}^{\langle\psi,1\rangle}$. Thus $\min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\left(b \otimes b_{\mathcal{U}}^{\langle\psi,1\rangle}\right) = c\left(b \oplus b_{\mathcal{U}}^{\langle\psi,1\rangle}\right) = b(\neg\psi)$, and in fact it is easy to see that $\max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_{\mathcal{U}}^{\langle\psi,1\rangle}(\varphi) - p^b(\varphi)\right) = b_{\mathcal{U}}^{\langle\psi,1\rangle}(\psi) - p^b(\psi) = 1 - (1 - b(\neg\psi)) = b(\neg\psi)$. In a moment we shall consider other special cases in which Conjecture 55 holds.

In general there are many JSSD's $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ which minimize the conflict $c(b_1 \otimes b_2)$, and a rule for constructing one of them from b_1 and b_2, if exists, is surely very complicated. If the beliefs considered are simple (i.e., the corresponding SSD's have few non-zero values), a JSSD which minimize the conflict can normally be constructed easily by reasoning on the particular situation. However, it can be interesting to define simple rules of combination such that the respective conflict is normally close to the minimum ("normally" is to be interpreted in an intuitive way, its meaning could be made more precise if we would consider certain kinds of beliefs as being "normal"), for instance the following one.

Definition 56 *Given two beliefs b_1 and b_2 on $\mathcal{U}$, for every $s \subset V_{\mathcal{U}}$ define*

$$\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s) = \min\left(m_{\mathcal{U}}^{b_1}(s), m_{\mathcal{U}}^{b_2}(s)\right)$$

and $m'_i(s) = m_{\mathcal{U}}^{b_i}(s) - \underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s)$ (for $i \in \{1, 2\}$); then for every $s, s' \subset V_{\mathcal{U}}$ such that $s \neq s'$ define $\underline{m}'(s, s') = \begin{cases} \frac{m'_1(s) m'_2(s')}{\max\left(\sum_{s'' \cap s \neq \emptyset} m'_2(s''), \sum_{s'' \cap s' \neq \emptyset} m'_1(s'')\right)} & \text{if } s \cap s' \neq \emptyset, \\ 0 & \text{if } s \cap s' = \emptyset, \end{cases}$ and

$$\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s') = \underline{m}'(s, s') + \frac{m''_1(s)\, m''_2(s')}{\sum_{s'' \subset V_{\mathcal{U}}} m''_1(s)},$$

where $m''_1(s) = m'_1(s) - \sum_{s' \neq s} \underline{m}'(s, s')$ and $m''_2(s) = m'_2(s) - \sum_{s' \neq s} \underline{m}'(s', s)$.

This definition can be interpreted as follows: the largest part possible of the support $m_{\mathcal{U}}^{b_1}(s)$ is delivered to $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s)$, a part of what remains is given to the $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s')$ such that $s \cap s' \neq \emptyset$ (this part is the biggest possible for a symmetrical and homogeneous rule), and only if something still remains, then it could be assigned also to the $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s')$ such that $s \cap s' = \emptyset$. This definition could seem recursive, but it is not: I have divided it in three parts only to clarify it a little; for this sake I have also written fractions whose denominator could be 0: in this case also the numerator would be 0, and $\frac{0}{0}$ is to be interpreted as 0. It is not difficult to see that $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}$ is a JSSD with marginal beliefs b_1 and b_2; the respective conflict of b_1 and b_2 is denoted by $c\,(b_1 \odot b_2)$, and the respective combination of b_1 and b_2 by $b_1 \odot b_2$. Only in the third part of the definition, some support could be assigned to the $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s')$ such that $s \cap s' = \emptyset$: therefore, intuitively, the conflict $c\,(b_1 \odot b_2) = \sum_{s \cap s' = \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s')$ will be low.

Example 57 *Consider the situation of Example 52: from the two radio stations I have obtained the beliefs $b_{\{r\}}^{\langle r,\alpha\rangle}$ and $b_{\{r\}}^{\langle r,\beta\rangle}$, where $\alpha \geq \beta$; if we use the rule $\odot$ to combine them, then we obtain the result $b_{\{r\}}^{\langle r,\alpha\rangle}$, which corresponds to the assumption of complete dependence (which is almost certainly the more sensible one). However, this example does not present any conflict between the beliefs; to introduce some conflict we can modify the situation in the following way. Assume that both weather forecasters have said that it is likely that tomorrow it will rain in Zürich, but the first one has been more precise (instead of being more convincing) and has added that nevertheless there is also a little probability that tomorrow it will not rain. We can model this new situation by considering $b_1 = b_{\{r\}}^{\{\langle r,\alpha\rangle, \langle \neg r,\beta\rangle\}}$ (where $\beta \leq 1 - \alpha$) and $b_2 = b_{\{r\}}^{\langle r,\alpha\rangle}$ as the beliefs that I have obtained from the first and the second radio stations, respectively. Using the rule $\odot$, we obtain the result $b_1 \odot b_2 = b_1$ (which is probably the more sensible one, since it is very likely that b_1 and b_2 are formulations of the same weather prediction, but b_1 is more precise than b_2) and there is no conflict between the two beliefs: $c\,(b_1 \odot b_2) = 0$ (thus in this case the rule $\odot$ minimize the conflict and Conjecture 55 is correct). Whereas using Dempster's rule, there is conflict between the two beliefs ($c\,(b_1 \oplus b_2) = \alpha\beta$) and we obtain a belief such that $(b_1 \oplus b_2)(r) = \frac{2\alpha - \alpha^2 - \alpha\beta}{1 - \alpha\beta} \geq \alpha = b_1(r)$ (the belief in r has increased, since r has been confirmed by a second source of information which is considered independent from the first one) and $(b_1 \oplus b_2)(\neg r) = \frac{\beta - \alpha\beta}{1 - \alpha\beta} \leq \beta = b_1(\neg r)$ (the belief in $\neg r$ has decreased, since $\neg r$ has been disconfirmed by a second source of information).*

In this example, the beliefs considered are similar, but in situations in which they are opposite, $c\,(b_1 \odot b_2)$ could be larger than $c\,(b_1 \oplus b_2)$ (in these situations, however, normally also the minimal conflict is large): for instance, if $\varphi \not\approx \top$, $c\left(b_{\mathcal{U}}^{\langle\varphi,\alpha\rangle} \odot b_{\mathcal{U}}^{\langle\neg\varphi,\alpha\rangle}\right) = \alpha$, whereas $c\left(b_{\mathcal{U}}^{\langle\varphi,\alpha\rangle} \oplus b_{\mathcal{U}}^{\langle\neg\varphi,\alpha\rangle}\right) = \alpha^2$ and the minimal conflict is $\max(2\alpha - 1, 0)$ (also in this case Conjecture 55 is valid). It is not difficult to think to rules of combination which are more complicated and behave better (i.e., lead to smaller conflicts) than the rule $\odot$. But this rule is relatively simple and, under the assumption that the beliefs to be combined are not too much contrasting, it could be very useful.

As I have said in Section 1.9, Dempster's rule of combination has some problem in dealing with epistemic probabilities; two epistemic probabilities b and b' on $\mathcal{U}$ are completely defined by the respective probabilities of the $n = 2^{|\mathcal{U}|}$ elementary events (if $V_{\mathcal{U}} = \{v_1, \ldots, v_n\}$, let $p_i = m_{\mathcal{U}}^{b}(\{v_i\})$ and $p'_i = m_{\mathcal{U}}^{b'}(\{v_i\})$). The combination of

b and b' with Dempster's rule would cause the conflict $c\,(b \oplus b') = 1 - \sum_{i=1}^{n} p_i p_i'$, which is in general very large; whereas the above defined rule leads to the minimal conflict $c\,(b \odot b') = \sum_{p_i > p_i'} (p_i - p_i') = \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b\,(\varphi) - p^{b'}(\varphi)\right) = \min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\,(b_1 \otimes b_2)$ (therefore, also in the case of two epistemic probabilities Conjecture 55 is valid). In particular, with Dempster's rule we would get the absurd conflict $c\,(b \oplus b) = 1 - \sum_{i=1}^{n} (p_i)^2$, whereas $c\,(b \odot b) = 0$.

From the work of Klawonn and Schwecke it follows that at least one of the three properties (commutativity, associativity and the invariance under vacuous extension), which characterize Dempster's rule of combination among the automatic rules which are describable through the JSSD's and satisfy the two technical assumptions, is not shared by the rule $\odot$ (since this satisfies the technical assumptions). It is easy to see that $\odot$ is commutative (since $\underline{m}_{\mathcal{U}}^{b_1 \odot b_2}(s, s') = \underline{m}_{\mathcal{U}}^{b_2 \odot b_1}(s', s)$) and invariant under vacuous extension; therefore it cannot be associative. The invariance under vacuous extension is a strictly necessary property, since the difference between a belief and one of its vacuous extensions is purely formal. The commutativity is positive in symmetrical situations in which the beliefs to be combined have the same importance, but in other situations we could prefer that one of the two beliefs would have a prominent role in the combination: since these situations can be precisely worked out with other methods (such as metabeliefs and discounting, which I shall consider later), the commutativity is surely welcome. I have already spoken about the associativity: I do not consider it necessary (even if certainly very practical), and I give it up with pleasure if in exchange I obtain a property which is surely much more desirable, for a rule to be applied without reflecting: the idempotency. In fact, $b \odot b = b$, whereas in general $b \oplus b \neq b$ and, as we have seen before, $c\,(b \oplus b)$ can be dreadfully close to 1.

2.3 Nonmonotonicity

Classical propositional logic is monotonic: if $\varphi \vdash \chi$, then $(\varphi \wedge \psi) \vdash \chi$. But this is true only if we allow inconsistent beliefs: in fact $(\varphi \wedge \psi)$ could be (completely) inconsistent, and so every proposition would be a conclusion, in particular χ. Since in the present formulation inconsistent beliefs are not allowed, even in the classical case monotonicity is not assured: given $b_{\mathcal{U}}^{\langle\varphi,1\rangle}$ and $b_{\mathcal{U}}^{\langle\psi,1\rangle}$, if $\{\langle(\varphi \wedge \psi), 1\rangle\}$ is inconsistent, then $b_{\mathcal{U}}^{\langle\varphi,1\rangle} \otimes b_{\mathcal{U}}^{\langle\psi,1\rangle}$ is not defined (since $c\left(b_{\mathcal{U}}^{\langle\varphi,1\rangle} \otimes b_{\mathcal{U}}^{\langle\psi,1\rangle}\right) = I_{\{\langle(\varphi\wedge\psi),1\rangle\}} = 1$, and therefore $\underline{m}_{\mathcal{U}}^{b_{\mathcal{U}}^{\langle\varphi,1\rangle} \otimes b_{\mathcal{U}}^{\langle\psi,1\rangle}}$ is not acceptable). In the particular case of classical beliefs, the inconsistency can only be complete, thus we have the following dichotomy: either there is monotonicity, or the combination is not defined. In the general case, the inconsistence can also be only partial and thus surmountable: this situation leads to nonmonotonicity.

Theorem 58 *Given two beliefs* b_1 *and* b_2 *on* $\mathcal{U}$*, an acceptable JSSD* $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ *and a formula* φ *on* $\mathcal{U}$*,*

$$\frac{\max\left(b_1\left(\varphi\right), b_2\left(\varphi\right)\right) - c\left(b_1 \otimes b_2\right)}{1 - c\left(b_1 \otimes b_2\right)} \leq \left(b_1 \otimes b_2\right)\left(\varphi\right) \leq p^{b_1 \otimes b_2}\left(\varphi\right) \leq \frac{\min\left(p^{b_1}\left(\varphi\right), p^{b_2}\left(\varphi\right)\right)}{1 - c\left(b_1 \otimes b_2\right)}.$$

Proof. From Proposition 35 follows that

$$\begin{aligned} p^{b_1 \otimes b_2}\left(\varphi\right) &= \sum_{s \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} m_{\mathcal{U}}^{b_1 \otimes b_2}\left(s\right) = \sum_{s \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} \tfrac{1}{1 - c(b_1 \otimes b_2)} \sum_{s' \cap s'' = s} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) = \\ &= \tfrac{1}{1 - c(b_1 \otimes b_2)} \sum_{s' \cap s'' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \leq \\ &\leq \tfrac{1}{1 - c(b_1 \otimes b_2)} \sum_{s' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} \sum_{s'' \subset V_{\mathcal{U}}} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) = \\ &= \tfrac{1}{1 - c(b_1 \otimes b_2)} \sum_{s' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} m_{\mathcal{U}}^{b_1}\left(s'\right) = \tfrac{1}{1 - c(b_1 \otimes b_2)} p^{b_1}\left(\varphi\right); \end{aligned}$$

analogously $p^{b_1 \otimes b_2}\left(\varphi\right) \leq \frac{1}{1 - c(b_1 \otimes b_2)} p^{b_2}\left(\varphi\right)$, and therefore $p^{b_1 \otimes b_2}\left(\varphi\right) \leq \frac{\min\left(p^{b_1}(\varphi), p^{b_2}(\varphi)\right)}{1 - c(b_1 \otimes b_2)}$.
From this we obtain also

$$\begin{aligned} \left(b_1 \otimes b_2\right)\left(\varphi\right) &= 1 - p^{b_1 \otimes b_2}\left(\neg\varphi\right) \geq 1 - \tfrac{\min\left(p^{b_1}(\neg\varphi), p^{b_2}(\neg\varphi)\right)}{1 - c(b_1 \otimes b_2)} = \\ &= \tfrac{\max\left(1 - p^{b_1}(\neg\varphi), 1 - p^{b_2}(\neg\varphi)\right) - c(b_1 \otimes b_2)}{1 - c(b_1 \otimes b_2)} = \tfrac{\max(b_1(\varphi), b_2(\varphi)) - c(b_1 \otimes b_2)}{1 - c(b_1 \otimes b_2)}; \end{aligned}$$

and from Theorem 36 the last inequality: $\left(b_1 \otimes b_2\right)\left(\varphi\right) \leq p^{b_1 \otimes b_2}\left(\varphi\right)$. ■

Corollary 59 *Given two sets of propositional symbols* $\mathcal{U}$ *and* $\mathcal{V}$*, two beliefs* b_1 *and* b_2 *on* $\mathcal{U}$*, an acceptable JSSD* $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ *and a grade* α*,*

- $\left(\mathcal{D}_{\mathcal{V}}\left(b_1\right) \cup \mathcal{D}_{\mathcal{V}}\left(b_2\right)\right) \subset \mathcal{D}_{\mathcal{V}}\left(b_1 \otimes b_2\right)$;
- $c\left(b_1 \otimes b_2\right) = 0 \;\Rightarrow\; \left(\mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1\right) \cup \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_2\right)\right) \subset \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1 \otimes b_2\right)$.

Proof. Since $\varphi \in \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1\right) \cup \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_2\right) \Rightarrow \max\left(b_1\left(\varphi\right), b_2\left(\varphi\right)\right) \geq \alpha$,

- $\varphi \in \mathcal{D}_{\mathcal{V}}\left(b_1\right) \cup \mathcal{D}_{\mathcal{V}}\left(b_2\right) \Rightarrow \left(b_1 \otimes b_2\right)\left(\varphi\right) \geq \frac{1 - c(b_1 \otimes b_2)}{1 - c(b_1 \otimes b_2)} = 1 \Rightarrow \varphi \in \mathcal{D}_{\mathcal{V}}\left(b_1 \otimes b_2\right)$;
- if $c\left(b_1 \otimes b_2\right) = 0$, then
  $\varphi \in \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1\right) \cup \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_2\right) \Rightarrow \left(b_1 \otimes b_2\right)\left(\varphi\right) \geq \frac{\alpha}{1} = \alpha \Rightarrow \varphi \in \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1 \otimes b_2\right)$. ■

Corollary 59 states that if the conflict is partial (i.e., less than 1), then we have certainly monotonicity for the deductive conclusions; therefore a deductive conclusion will always be maintained, unless we try to combine two totally conflicting beliefs. This is not valid for inductive conclusions: if $\alpha < 1$, then in general $\mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1\right) \not\subset \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1 \otimes b_2\right)$ (nonmonotonicity), and even $\left(\mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1\right) \cap \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_2\right)\right) \not\subset \mathcal{C}_{\mathcal{V}}^{\alpha}\left(b_1 \otimes b_2\right)$. The lower bound for $\left(b_1 \otimes b_2\right)\left(\varphi\right)$

given in Theorem 58 is the best one that we can obtain without considering the particular situations (to be precise, the best lower bound is the maximum of this one and 0): this points out the negative effect of the conflict regarding monotonicity. In fact, as Corollary 59 states, if there is no conflict, then we have monotonicity also for the inductive conclusions. In this context Conjecture 55 can be read as follows: if the monotonicity is acceptable, then it is feasible; in other words, if there is no $\varphi \in \mathcal{L}_{\mathcal{U}}$ such that $b_1(\varphi) > p^{b_2}(\varphi)$ (remember that $b_2(\varphi) > p^{b_1}(\varphi) \Rightarrow b_1(\neg\varphi) > p^{b_2}(\neg\varphi)$), then there is a JSSD without conflict. The monotonicity for strong conclusions is not assured even if there is no conflict: if $\alpha < 1$, then in general $(\mathcal{S}_{\mathcal{V}}^{\alpha}(b_1) \cap \mathcal{S}_{\mathcal{V}}^{\alpha}(b_2)) \not\subset \mathcal{S}_{\mathcal{V}}^{\alpha}(b_1 \otimes b_2)$, even if $c(b_1 \otimes b_2) = 0$ (even if φ is a strong conclusion of b_1 and of b_2, their combination can bring some support to $\neg\varphi$).

Theorem 60 *Given two beliefs b_1 and b_2 such that $c(b_1 \oplus b_2) < 1$, and a formula φ on $\mathcal{U}$,*

$$\frac{b_1(\varphi) + b_2(\varphi) - b_1(\varphi)\, b_2(\varphi) - c(b_1 \oplus b_2)}{1 - c(b_1 \oplus b_2)} \leq (b_1 \oplus b_2)(\varphi) \leq p^{b_1 \oplus b_2}(\varphi) \leq \frac{p^{b_1}(\varphi)\, p^{b_2}(\varphi)}{1 - c(b_1 \oplus b_2)}.$$

Proof. As in the proof of Theorem 58, we obtain $(b_1 \oplus b_2)(\varphi) \leq p^{b_1 \oplus b_2}(\varphi)$ and

$$\begin{aligned} p^{b_1 \oplus b_2}(\varphi) &= \tfrac{1}{1-c(b_1 \oplus b_2)} \sum_{s' \cap s'' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \oplus b_2}(s', s'') = \\ &= \tfrac{1}{1-c(b_1 \oplus b_2)} \sum_{s' \cap s'' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} m_{\mathcal{U}}^{b_1}(s')\, m_{\mathcal{U}}^{b_2}(s'') \leq \\ &\leq \tfrac{1}{1-c(b_1 \oplus b_2)} \left(\sum_{s' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} m_{\mathcal{U}}^{b_1}(s') \right) \left(\sum_{s'' \cap s_{\mathcal{U}}(\varphi) \neq \emptyset} m_{\mathcal{U}}^{b_2}(s'') \right) = \tfrac{p^{b_1}(\varphi) p^{b_2}(\varphi)}{1-c(b_1 \oplus b_2)}. \end{aligned}$$

From this follows also

$$\begin{aligned} (b_1 \oplus b_2)(\varphi) &= 1 - p^{b_1 \oplus b_2}(\neg\varphi) \geq 1 - \tfrac{p^{b_1}(\neg\varphi) p^{b_2}(\neg\varphi)}{1-c(b_1 \oplus b_2)} = \\ &= \tfrac{1-(1-b_1(\varphi))(1-b_2(\varphi))-c(b_1 \oplus b_2)}{1-c(b_1 \oplus b_2)} = \tfrac{b_1(\varphi)+b_2(\varphi)-b_1(\varphi)b_2(\varphi)-c(b_1 \oplus b_2)}{1-c(b_1 \oplus b_2)}. \blacksquare \end{aligned}$$

Corollary 61 *Given two sets of propositional symbols $\mathcal{U}$ and $\mathcal{V}$, two beliefs b_1 and b_2 on $\mathcal{U}$ such that $c(b_1 \oplus b_2) < 1$, and a grade α,*

- $\alpha \geq c(b_1 \oplus b_2) \;\Rightarrow\; (\mathcal{C}_{\mathcal{V}}^{\alpha}(b_1) \cap \mathcal{C}_{\mathcal{V}}^{\alpha}(b_2)) \subset \mathcal{C}_{\mathcal{V}}^{\alpha}(b_1 \oplus b_2)$.

Proof. If $\alpha \geq c(b_1 \oplus b_2)$, then

$$\begin{aligned} \varphi \in (\mathcal{C}_{\mathcal{V}}^{\alpha}(b_1) \cap \mathcal{C}_{\mathcal{V}}^{\alpha}(b_2)) \Rightarrow (b_1 \oplus b_2)(\varphi) &\geq \tfrac{b_1(\varphi)+b_2(\varphi)-b_1(\varphi)b_2(\varphi)-c(b_1 \oplus b_2)}{1-c(b_1 \oplus b_2)} = \\ &= \tfrac{b_1(\varphi)(1-c(b_1 \oplus b_2))+(b_2(\varphi)-c(b_1 \oplus b_2))(1-b_1(\varphi))}{1-c(b_1 \oplus b_2)} \geq \\ &\geq \tfrac{b_1(\varphi)(1-c(b_1 \oplus b_2))}{1-c(b_1 \oplus b_2)} = b_1(\varphi) \geq \alpha \Rightarrow \\ \Rightarrow \varphi \in \mathcal{C}_{\mathcal{V}}^{\alpha}(b_1 \oplus b_2). \blacksquare \end{aligned}$$

Since Dempster's rule of combination has a mathematically simple form, we can obtain a lower bound for $(b_1 \oplus b_2)(\varphi)$ more precise than the one given in Theorem 58; as this one, the lower bound of Theorem 60 is the best one that we can obtain without considering the particular situations (precisely, the best lower bound is the maximum of this one and 0). This allows the result of Corollary 61: a weak monotonicity for the inductive conclusions with a grade not smaller than the conflict (where "weak" means that in general $\mathcal{C}_{\mathcal{V}}^{\alpha}(b_1) \not\subseteq \mathcal{C}_{\mathcal{V}}^{\alpha}(b_1 \oplus b_2)$ and $\mathcal{C}_{\mathcal{V}}^{\alpha}(b_2) \not\subseteq \mathcal{C}_{\mathcal{V}}^{\alpha}(b_1 \oplus b_2)$).

Example 62 *In Example 57 we had two beliefs on $\{r\}$: $b_1 = b_{\{r\}}^{\{\langle r,\alpha\rangle,\langle\neg r,\beta\rangle\}}$ and $b_2 = b_{\{r\}}^{\langle r,\alpha\rangle}$. Using the rule of combination $\odot$, the lower bound of Theorem 58 is attained for every formula φ on $\{r\}$: $(b_1 \odot b_2)(\varphi) = \max(b_1(\varphi), b_2(\varphi)) = b_1(\varphi)$; as Corollary 59 states, since there is no conflict, we have monotonicity for the conclusions. Using Dempster's rule, the lower bound of Theorem 60 is attained for every formula φ on $\{r\}$ such that $\varphi \not\approx \bot$: $(b_1 \oplus b_2)(\varphi) = \frac{b_1(\varphi)+b_2(\varphi)-b_1(\varphi)b_2(\varphi)-c(b_1\oplus b_2)}{1-c(b_1\oplus b_2)}$, since $(b_1 \oplus b_2)(r) = \frac{\alpha+\alpha-\alpha^2-\alpha\beta}{1-\alpha\beta}$, $(b_1 \oplus b_2)(\neg r) = \frac{\beta+0-0-\alpha\beta}{1-\alpha\beta}$ and $(b_1 \oplus b_2)(\top) = \frac{1+1-1-\alpha\beta}{1-\alpha\beta}$. Since there is the conflict $c(b_1 \oplus b_2) = \alpha\beta$, we have nonmonotonicity for the conclusions: $\neg r \in \mathcal{C}_{\{r\}}^{\beta}(b_1)$, but $\neg r \notin \mathcal{C}_{\{r\}}^{\beta}(b_1 \oplus b_2)$; as Corollary 61 states, since $\alpha \geq \alpha\beta$, we have a weak nonmonotonicity for the conclusions with grade α: $r \in \mathcal{C}_{\{r\}}^{\alpha}(b_1) \cap \mathcal{C}_{\{r\}}^{\alpha}(b_2)$ and $r \in \mathcal{C}_{\{r\}}^{\alpha}(b_1 \oplus b_2)$. In Example 52 (where there is no conflict), the lower bounds of Theorems 58 and 60 are attained (for every formula) in the cases $\gamma = \frac{\beta}{\alpha}$ (extreme dependence) and $\gamma = \beta$ (independence), respectively: this stresses the quality of these lower bounds.*

Example 63 *An interesting example of nonmonotonic reasoning is the following one: Mr. and Mrs. Milton live in New York and would like to spend their holidays in a natural environment; Mrs. Milton has read in a touristic magazine about the wonderful landscapes of the Dolomites and suggests to go there. Her husband thinks that it is a good idea, but later he notices that the same touristic magazine describes also a region of Oregon where almost every day the sun shines; since he is persuaded that in Oregon it always rains, he concludes that the magazine is unreliable and he does not want to go to the Dolomites anymore. But the following day he meets a friend who tells him that he has been in a place in Oregon where it rains hardly ever; therefore Mr. Milton recovers his confidence in the magazine and accepts the proposal of his wife. We can model the reasoning of Mr. Milton by considering the propositional symbols m, d and o, which represent the propositions "The touristic magazine is reliable", "The Dolomites are a beautiful place" and "There is a region of Oregon where it rains hardly ever", respectively (let $\mathcal{U} = \{m, d, o\}$). The relation between m and d can be described with the belief $b_1 = b_{\mathcal{U}}^{\{\langle(m\leftrightarrow d),\alpha\rangle,\langle(m\rightarrow d),1-\alpha\rangle\}}$: m implies certainly d (if the magazine is reliable, then the Dolomites are beautiful), but also d gives some support to m (if the Dolomites are beautiful, then the belief in the reliability of the magazine increases); analogously $b_2 = b_{\mathcal{U}}^{\{\langle(m\leftrightarrow o),\alpha\rangle,\langle(m\rightarrow o),1-\alpha\rangle\}}$ describes the relation between m and o. These two beliefs can be combined using Dempster's rule, not because* *they come from two independent sources of information, but because the result of this combination is sensible: both the certainties of d and o would augment the belief in m by a portion α of the maximal increase possible, thus the certainty of $(d \wedge o)$ would cause the belief $\alpha + \alpha(1-\alpha)$ in m (in fact $(b_1 \oplus b_2)(((d \wedge o) \rightarrow m)) = 2\alpha - \alpha^2$). Since it is very difficult to distinguish the sources of the two beliefs, from the conceptual standpoint their combination does not make any sense: we should define a single belief on $\mathcal{U}$ which includes both pieces of information; on the grounds of what is stated above, I think that $b_1 \oplus b_2$ is a good choice for such a belief, thus only the value of α remains to be chosen: for instance set $\alpha = 0.6$. At first Mr. Milton gave some (a priori) support to the reliability of the magazine; we can describe this with the simple support β of m: $b_3 = b_{\mathcal{U}}^{\langle m,\beta\rangle}$, for instance set $\beta = 0.7$. Since he was persuaded that in Oregon it always rains, he supported also $\neg o$: $b_4 = b_{\mathcal{U}}^{\langle \neg o,\gamma\rangle}$; since $\neg o$ is a lot more precise than "In Oregon it always rains", γ should not be so large as the support that he would have given to this sentence: for instance set $\gamma = 0.8$ (we could introduce a propositional symbol o' to represent "In Oregon it always rains" and describe the relation between o and o' with another belief, but this would complicate the model). The last belief that we have to consider is the one that Mr. Milton gets from his friend: the simple support δ of o: $b_5 = b_{\mathcal{U}}^{\langle o,\delta\rangle}$; where δ should be large, since the friend is reliable and has been in Oregon: for instance set $\delta = 0.95$. I think that these beliefs can easily be seen as coming from independent sources of information, thus I combine them using Dempster's rule. Therefore the reasoning of Mr. Milton can be described as follows: at first he had the belief $b_3 \oplus b_4$, then, when his wife shows him the description of the Dolomites, he has the belief $b_1 \oplus b_3 \oplus b_4$, and when he notice the article about the region of Oregon where it never rains, he combines this belief with b_2 obtaining $b_1 \oplus b_2 \oplus b_3 \oplus b_4$; finally, after he has met the friend, he has the belief $b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5$. The following table shows the evolution of his beliefs about d, $\neg d$, o, $\neg o$, m and $\neg m$.*

	$b_3 \oplus b_4$	$b_1 \oplus b_3 \oplus b_4$	$b_1 \oplus b_2 \oplus b_3 \oplus b_4$	$b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5$
d	0	β	$\frac{\beta(1-\gamma)}{1-\beta\gamma}$	$\frac{(1-\gamma)(\alpha\delta+\beta-\alpha\beta\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$
$\neg d$	0	0	$\frac{\alpha(1-\beta)\gamma}{1-\beta\gamma}$	$\frac{\alpha(1-\beta)\gamma(1-\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$
o	0	0	$\frac{\beta(1-\gamma)}{1-\beta\gamma}$	$\frac{(1-\gamma)(\beta+\delta-\beta\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$
$\neg o$	γ	γ	$\frac{(1-\beta)\gamma}{1-\beta\gamma}$	$\frac{(1-\beta)\gamma(1-\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$
m	β	β	$\frac{\beta(1-\gamma)}{1-\beta\gamma}$	$\frac{(1-\gamma)(\alpha\delta+\beta-\alpha\beta\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$
$\neg m$	0	0	$\frac{(1-\beta)\gamma}{1-\beta\gamma}$	$\frac{(1-\beta)\gamma(1-\delta)}{1-\gamma(\beta+\delta-\beta\delta)}$

With the proposed values for α, β, γ and δ, the (approximate) results are shown in the *following table.*

	$b_3 \oplus b_4$	$b_1 \oplus b_3 \oplus b_4$	$b_1 \oplus b_2 \oplus b_3 \oplus b_4$	$b_1 \oplus b_2 \oplus b_3 \oplus b_4 \oplus b_5$
d	0	0.70	0.32	0.82
$\neg d$	0	0	0.33	0.03
o	0	0	0.32	0.93
$\neg o$	0.80	0.80	0.55	0.06
m	0.70	0.70	0.32	0.82
$\neg m$	0	0	0.55	0.06

I think that these results are easily interpretable as a description of the evolution of Mr. Milton's beliefs; apart from the one in o, the other five beliefs present a nonmonotonic evolution: this is due to the conflicts between b_2 and $b_1 \oplus b_3 \oplus b_4$, and between the combination of these two beliefs and b_5.

$$c\,(b_2 \oplus (b_1 \oplus b_3 \oplus b_4)) = \beta\gamma \text{ and } c\,((b_1 \oplus b_2 \oplus b_3 \oplus b_4) \oplus b_5) = \frac{(1-\beta)\,\gamma\delta}{1-\beta\gamma};$$

with the proposed values for the variables these two conflicts are 0.56 *and (approximately)* 0.52*, respectively. Of course the model presented for Mr. Milton's beliefs is not the only one possible; but I consider it quite complete, in spite of its simplicity. However, even if we choose this model, the values of the variables α, β, γ and δ are arbitrary: in the first table I have reported the general expressions, so that whoever wants to, can easily calculate the results with his/her own choices for these values.*

2.4 Metabeliefs

If we combine two beliefs, which have been obtained from two sources of information, this means that we believe both sources to be completely reliable. Thus, we have a belief about the reliability of the sources of information: I shall call such a belief a "metabelief" (in practice a source of information is often identified with the belief we get from it, therefore a metabelief is a belief about the reliability of others beliefs: this explains the name). Up to now I have (implicitly) considered only metabeliefs which are certainties: the certainties that one, two or more sources of information are reliable; from now on, on the contrary, I shall accept any kind of beliefs as metabeliefs. Given n sources of information which give us the beliefs $b_1, \ldots, b_n$, respectively, on a topic $\mathcal{U}$, a metabelief about the reliability of these sources can be described with a belief b on $\mathcal{B} = \{b_1, \ldots, b_n\}$: $b\,(b_i)$ expresses the belief about the reliability of the i-th source of information. With this very practical definition, $b_1, \ldots, b_n$ are both propositional symbols of the object-language $\mathcal{L}_\mathcal{B}$ and symbols of the metalanguage which indicate n beliefs on $\mathcal{U}$: I think that this will not lead to confusions, if we set the obvious condition $\mathcal{U} \cap \mathcal{B} = \emptyset$ (i.e., the propositional symbols $b_1, \ldots, b_n$ are not arguments of the beliefs $b_1, \ldots, b_n$).

Of course, to the metabelief $b_{\mathcal{B}}^{\langle b_1,1\rangle}$ (the certainty of b_1) corresponds the belief b_1 on $\mathcal{U}$; whereas to $b_{\mathcal{B}}^{\langle (b_1\wedge b_2),1\rangle}$ corresponds a combination of b_1 and b_2. In general, given a metabelief b on $\mathcal{B}$, we would like to define the corresponding belief on $\mathcal{U}$; this purpose can be achieved by defining a conditional belief $b \mid \mathcal{B}$ on $\mathcal{U}$: the belief on $\mathcal{U}$ corresponding to the metabelief b will be

$$(b \otimes (b \mid \mathcal{B}))\downarrow_{\mathcal{U}} \tag{3}$$

(where $b \otimes (b \mid \mathcal{B})$ is the notation without vacuous extensions introduced in Section 2.2). Of course, the conditional belief $b \mid \mathcal{B}$ should satisfy some obvious requirements: first of all, that $b \mid b_i = b_i$ for every $i \in \{1,\ldots,n\}$ (if all we know is that a source of information is reliable, then we should accept the respective belief). If all we know is that a source of information is unreliable, then we should not accept any belief coming from the sources: this can be expressed through the condition $b \mid \neg b_i = b_{\mathcal{U}}^{\emptyset}$. Therefore, I think it is sensible to require that $b \mid \mathcal{B}$ satisfies the following conditions (for every $i \in \{1,\ldots,n\}$, $\varphi \in \mathcal{L}_{\mathcal{B}}$ and $B \subset \{\neg b_1,\ldots,\neg b_n\}$):

- $b \mid b_i = b_i$;
- if $(\varphi \wedge \bigvee B) \nsim \bot$, then $b \mid (\varphi \wedge \bigvee B) = b \mid \varphi$;
- $b \mid (\varphi \vee \bigwedge B) = b_{\mathcal{U}}^{\emptyset}$;
- $b \mid \bot = b_{\mathcal{U}}^{\emptyset}$.

It can be easily proven that these conditions are not contradictory: for every n there is a conditional belief $b \mid \mathcal{B}$ on $\mathcal{U}$ which satisfies these requirements. The first condition is clear, whereas the second one can be interpreted as follows: the knowledge of the unreliability of some sources of information does not change our belief on $\mathcal{U}$, unless this one was based on these sources. For instance, this requirement implies that, for $i \neq j$, $b \mid (b_i \wedge \neg b_j) = b_i$ and $b \mid (b_i \wedge (\neg b_i \vee \neg b_j)) = b_i$ (this is sensible, since $(b_i \wedge (\neg b_i \vee \neg b_j)) \sim (b_i \wedge \neg b_j)$). The third condition states that if it is possible that all the sources of information are unreliable, then our belief should be vacuous; in fact it is not difficult to prove that if the metabelief b is the certainty of ψ, then $p^b\left(\bigwedge\{\neg b_1,\ldots,\neg b_n\}\right) > 0$ (and thus $p^b\left(\bigwedge\{\neg b_1,\ldots,\neg b_n\}\right) = 1$) if and only if ψ can be written as $(\varphi \vee \bigwedge B)$. The last condition is formal, in the sense that, since the metabelief b (like any other belief) has to be consistent, surely $b(\bot) = 0$, and so it has no importance which belief we choose as $b \mid \bot$. I have set the requirement $b \mid \bot = b_{\mathcal{U}}^{\emptyset}$ solely because, for every n, the other three conditions permit only few choices about $b \mid \mathcal{B}$; and I do not want that the merely formal choice of $b \mid \bot$ could seem one of these.

Smets in [33] has developed a formalism for the combination of beliefs which has some similarities with the present one based on metabeliefs. Smets' formalism is a generalization of Dempster's rule (which Smets calls conjunctive: we are sure that both sources of information are reliable) and Smets' disjunctive rule for the combination of beliefs in the case in which we are only certain that at least one of the two sources of information is reliable (this rule was already proposed by Smets in his doctoral dissertation; see for instance [32]). The disjunctive combination of two beliefs b_1 and b_2 on $\mathcal{U}$ (which I shall denote with $b_1 \uplus b_2$), is constructed as follows: the support $m_{\mathcal{U}}^{b_1}(s)\, m_{\mathcal{U}}^{b_2}(s')$, instead of being assigned to $s \cap s'$ as in Dempster's rule, is assigned to $s \cup s'$. For instance $b_{\mathcal{U}}^{\langle\varphi,\alpha\rangle} \uplus b_{\mathcal{U}}^{\langle\psi,\beta\rangle} = b_{\mathcal{U}}^{\langle(\varphi\vee\psi),\alpha\beta\rangle}$; in general it can be shown that $(b_1 \uplus b_2)(\varphi) = b_1(\varphi)\, b_2(\varphi)$ (the disjunctive combination of two beliefs is their product). The difference between Smets' formalism and the present one is in the interpretation of the case of an unreliable source of information: according to Smets, if from a source we get the belief b and we consider this source as being unreliable, then we should assume a belief b' which is in a certain sense the contrary of b. I think that assuming the vacuous belief is much more in the spirit of the Dempster-Shafer theory, in which not believing in something does not mean believing its contrary.

Given a metabelief b on $\mathcal{B}$, there is another way to construct the corresponding belief on $\mathcal{U}$, besides the one based on the conditional belief $b \mid \mathcal{B}$. This new way is based on the notion of hypothetical belief, which can be seen as a simpler alternative to the notion of conditional belief; from the standpoint of the results, this new way is a special case of the already formulated one.

Definition 64 *Given a belief b on $\mathcal{U}$ and a propositional symbol $h \notin \mathcal{U}$, the hypothetical belief with hypothesis h and thesis b is the belief $h \rightharpoonup b$ on $\mathcal{U} \cup \{h\}$, such that*

$$h \rightharpoonup b = b_{\mathcal{U}\cup\{h\}}^{\{\langle(h\rightarrow\varphi_i),\alpha_i\rangle \mid i\in I\}}, \text{ where } b = b_{\mathcal{U}}^{\{\langle\varphi_i,\alpha_i\rangle \mid i\in I\}}.$$

It is easy to see that $h \rightharpoonup b$ is well-defined (i.e., it does not depend on the choice of the support distribution $\{\langle\varphi_i,\alpha_i\rangle \mid i \in I\}$ on $\mathcal{U}$), and that $(h \rightharpoonup b)((h \rightarrow \varphi)) = b(\varphi)$ for every $\varphi \in \mathcal{L}_{\mathcal{U}}$. The relations $\left((h \rightharpoonup b) \oplus b_{\mathcal{U}\cup\{h\}}^{\langle h,1\rangle}\right)\downarrow_{\mathcal{U}} = b$, $\left((h \rightharpoonup b) \oplus b_{\mathcal{U}\cup\{h\}}^{\langle\neg h,1\rangle}\right)\downarrow_{\mathcal{U}} = b_{\mathcal{U}}^{\emptyset}$ and $\left((h \rightharpoonup b) \oplus b_{\mathcal{U}\cup\{h\}}^{\emptyset}\right)\downarrow_{\mathcal{U}} = b_{\mathcal{U}}^{\emptyset}$ explain the name "hypothetical belief": if we are sure of the hypothesis h, then we obtain the thesis b; and in the two other extreme cases (if we are sure that the hypothesis is false or if we do not know anything about the truth of the hypothesis) we obtain the vacuous belief. Given a metabelief b on $\mathcal{B}$, we can consider the belief

$$\left((b\uparrow_{\mathcal{U}\cup\mathcal{B}}) \oplus ((b_1 \rightharpoonup b_1)\uparrow_{\mathcal{U}\cup\mathcal{B}}) \oplus \cdots \oplus ((b_n \rightharpoonup b_n)\uparrow_{\mathcal{U}\cup\mathcal{B}})\right)\downarrow_{\mathcal{U}} \tag{4}$$

as being the corresponding belief on $\mathcal{U}$. It is important to notice that all these n applications of Dempster's rule do not present any conflict (independently of the order in which the combinations are performed), and that the use of the rule $\odot$ instead of $\otimes$ would give exactly the same result (independently of the order in which this rule, which in general is not associative, is applied). These two considerations and the acceptability of the assumption of independence (since the hypothetical beliefs $(b_i \rightharpoonup b_i)$ are merely technical constructs) justify the use of Dempster's rule. To simplify the notation I shall write $b \oplus \bigoplus_{i=1}^{n} (b_i \rightharpoonup b_i)$ instead of $(b{\uparrow_{\mathcal{U}\cup\mathcal{B}}}) \oplus ((b_1 \rightharpoonup b_1){\uparrow_{\mathcal{U}\cup\mathcal{B}}}) \oplus \cdots \oplus ((b_n \rightharpoonup b_n){\uparrow_{\mathcal{U}\cup\mathcal{B}}})$, forgetting the vacuous extensions; thus (4) will be written as $\left(b \oplus \bigoplus_{i=1}^{n} (b_i \rightharpoonup b_i)\right){\downarrow_{\mathcal{U}}}$.

Consider now the case in which we have only one source of information (i.e., $n = 1$), which gives us the belief b_1 on a topic $\mathcal{U}$, and a metabelief b about the reliability of this source (i.e., a belief b on $\mathcal{B} = \{b_1\}$). In this case the conditional belief $b \mid \mathcal{B}$ on $\mathcal{U}$ is completely defined from the four conditions stated above: we must set $b \mid \bot = b_{\mathcal{U}}^{\emptyset}$, $b \mid b_1 = b_1$, $b \mid \neg b_1 = b \mid (\varphi \vee \bigwedge B) = b_{\mathcal{U}}^{\emptyset}$ (for instance if $\varphi \equiv \bot$ and $B = \{\neg b_1\}$) and $b \mid \top = b \mid (\varphi \vee \bigwedge B) = b_{\mathcal{U}}^{\emptyset}$ (for instance if $\varphi \equiv \top$ and $B = \emptyset$). Therefore the belief on $\mathcal{U}$ corresponding to the metabelief b is (3): $(b \otimes (b \mid \mathcal{B})){\downarrow_{\mathcal{U}}} = b\,(b_1)\, b_1 + (1 - b\,(b_1))\, b_{\mathcal{U}}^{\emptyset}$; that is, a weighted average of b_1 and the vacuous belief, where the weight of b_1 is the belief $b\,(b_1)$ in the reliability of the source of information. Using (4) we obtain the belief $(b \oplus (b_1 \rightharpoonup b_1)){\downarrow_{\mathcal{U}}} = b\,(b_1)\, b_1 + (1 - b\,(b_1))\, b_{\mathcal{U}}^{\emptyset}$; in this case the two methods lead to the same result.

In the Dempster-Shafer theory, the process leading from a belief b on $\mathcal{U}$ to the belief $\gamma b + (1 - \gamma)\, b_{\mathcal{U}}^{\emptyset}$ is called "discounting" with "discount rate" $1 - \gamma$ (for simplicity I shall write γb instead of $\gamma b + (1 - \gamma)\, b_{\mathcal{U}}^{\emptyset}$ to indicate the result of the discounting). From the above considerations it follows that discounting a belief with rate $1 - \gamma$ means assigning a belief γ to the reliability of the source of information from which we got the belief. Obviously (for $\gamma \neq 0$), if $\alpha \in [0, \gamma]$, then $\mathcal{C}_{\mathcal{V}}^{\alpha}(\gamma b) = \mathcal{C}_{\mathcal{V}}^{\alpha/\gamma}(b)$, and if $\alpha \in (\gamma, 1]$, then $\mathcal{C}_{\mathcal{V}}^{\alpha}(\gamma b) = \mathcal{T}_{\mathcal{V}}$; that is, a source of information cannot lead to non-tautological conclusions with a grade larger than its reliability, and in particular a not completely reliable source of information cannot give non-tautological certainties. The discounting of beliefs reduces the conflict when we combine them. Of course this statement has a clear sense only for well-defined rules of combination, for instance for Dempster's one: it is easy to see that $c\,(\gamma_1 b_1 \oplus \gamma_2 b_2) = \gamma_1 \gamma_2 c\,(b_1 \oplus b_2)$. If Conjecture 55 is correct, then $\min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\,(\gamma b_1 \otimes \gamma b_2) = \max\left(\gamma \min_{\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}} c\,(b_1 \otimes b_2) - (1 - \gamma)\,, 0\right)$; since, for every $\varphi \in \mathcal{L}_{\mathcal{U}} \setminus \mathcal{T}_{\mathcal{U}}$, $(\gamma b_1)(\varphi) - p^{\gamma b_2}(\varphi) = \gamma b_1(\varphi) - 1 + \gamma\left(1 - p^{b_2}(\varphi)\right) = \gamma\left(b_1(\varphi) - p^{b_2}(\varphi)\right) - (1 - \gamma)$, whereas $(\gamma b_1)(\top) - p^{\gamma b_2}(\top) = b_1(\top) - p^{b_2}(\top) = 0$. Therefore, by discounting the beliefs with rate $1 - \gamma$, the minimal conflict would surely not be larger than $2\gamma - 1$ (if this is bigger than 0, of course). Since the discounting of beliefs reduces the conflict, it has been proposed as a merely technical method to improve the monotonicity of the (artificial) reasoning based on Dempster-Shafer theory (see for instance [34]): given two beliefs b_1 and b_2, discount them (or one of them) with a rate depending on $c\,(b_1 \oplus b_2)$, and then combine them with Dempster's rule. From the technical standpoint these ideas are interesting, but from the conceptual one they are debatable, because, even if b_1 and b_2 come from independent sources of information, the discounted beliefs are certainly no more independent (since the discount rate depends on $c\,(b_1 \oplus b_2)$).

Consider now the case in which we have two sources of information (i.e., $n = 2$), which give us the beliefs b_1 and b_2, respectively, on a topic $\mathcal{U}$, and a metabelief about the reliability of these sources (i.e., a belief b on $\mathcal{B} = \{b_1, b_2\}$). It is easy to see that in this case the conditional belief $b \mid \mathcal{B}$ on $\mathcal{U}$ is defined from the four conditions stated above, except the $b \mid \varphi$ for $\varphi \sim (b_1 \wedge b_2)$ or $\varphi \sim (b_1 \vee b_2)$, which must be chosen. To define $b \mid (b_1 \wedge b_2)$, we have to choose a JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ (then we can set $b \mid (b_1 \wedge b_2) = b_1 \otimes b_2$); whereas $b \mid (b_1 \vee b_2)$ could be defined in a lot of different ways: for instance we could cautiously set $b \mid (b_1 \vee b_2) = b_{\mathcal{U}}^{\emptyset}$, or we could consider the disjunctive combination and set $b \mid (b_1 \vee b_2) = b_1 \uplus b_2$, or even, if b_1 is stronger than b_2 (i.e., $b_1 \geq b_2$), set $b \mid (b_1 \vee b_2) = b_2$. The belief (3) on $\mathcal{U}$ corresponding to the metabelief b is

$$\alpha\,(b \mid (b_1 \wedge b_2)) + \beta\,(b \mid b_1) + \gamma\,(b \mid b_2) + \delta\,(b \mid (b_1 \vee b_2)) + \varepsilon b_{\mathcal{U}}^{\emptyset}, \text{ where}$$

$$\alpha = b\,((b_1 \wedge b_2))\,,\ \beta = b\,(b_1) - \alpha,\ \gamma = b\,(b_2) - \alpha,$$
$$\delta = b\,((b_1 \vee b_2)) - \beta - \gamma - \alpha, \text{ and } \varepsilon = 1 - b\,((b_1 \vee b_2))\,.$$

Since $\alpha+\beta+\gamma+\delta+\varepsilon = \alpha+\beta+\gamma+(b\,((b_1 \vee b_2)) - \beta - \gamma - \alpha)+(1 - b\,((b_1 \vee b_2))) = 1$, the belief on $\mathcal{U}$ corresponding to the metabelief b is a (in my opinion very sensible) weighted average of the beliefs $b \mid (b_1 \wedge b_2)$, $b \mid b_1$, $b \mid b_2$, $b \mid (b_1 \vee b_2)$ and of the vacuous belief. It can be easily shown that using (4) we obtain the belief $\left(b \oplus \bigoplus_{i=1}^{2} (b_i \rightharpoonup b_i)\right) \downarrow_{\mathcal{U}} = (b \otimes (b' \mid \mathcal{B})) \downarrow_{\mathcal{U}}$, where $b' \mid \mathcal{B}$ is the conditional belief on $\mathcal{U}$ with the choices $b' \mid (b_1 \wedge b_2) = b_1 \oplus b_2$ and $b' \mid (b_1 \vee b_2) = b_1 \uplus b_2$. In this case the second method is equivalent to the special case of the first in which the two possible choices are those based on Dempster's rule and the disjunctive one, respectively; this gives a justification of the disjunctive rule based on Dempster's one (since this is used in the combinations of the second method).

A particular case of the situation with two beliefs is (seldom) considered in the Dempster-Shafer theory: the weighted average $\beta b_1 + (1 - \beta)\,b_2$ of two beliefs b_1 and b_2. This case is not clearly describable in terms of the belief in the reliability of the sources of information: to obtain a weighted average of b_1 and b_2, the metabelief b should have the properties $p^b\,((b_1 \wedge b_2)) = 0$ (which is clearly interpretable: we believe that the simultaneous reliability of the two sources is impossible) and $b\,(b_1) + b\,(b_2) = 1$ (which has no clear interpretation). However, the weighted average can be used to combine two beliefs: the weights have to be decided, and for this purpose a measure of information for the beliefs can be very useful (the measures of information are the topic of the next Section). Obviously, $(\mathcal{C}_{\mathcal{V}}^{\alpha}\,(b_1) \cap \mathcal{C}_{\mathcal{V}}^{\alpha}\,(b_2)) \subset \mathcal{C}_{\mathcal{V}}^{\alpha}\,(\beta b_1 + (1 - \beta)\,b_2)$: we have a weak monotonicity for the inductive conclusions of any grade; but in general (for $\beta \neq 1$) $\mathcal{D}_{\mathcal{V}}\,(b_1) \not\subset \mathcal{D}_{\mathcal{V}}\,(\beta b_1 + (1 - \beta)\,b_2)$: we do not have a (strong) monotonicity not even for the deductive conclusions; finally, of course this combinations are idempotent: $\beta b_1 + (1 - \beta)\,b_1 = b_1$.

The cases with $n > 2$ sources of information are analogous: the results of both methods are weighted averages of beliefs, where the weights are fixed from the metabelief in a clear and sensible way. The second method (the one based on hypothetical beliefs) is always equivalent to the special case of the first method in which the possible choices for the conditional beliefs are the standard ones of the Dempster-Shafer theory (i.e., the choices based on Dempster's rule and the disjunctive one).

2.5 Measures of information

So many authors have introduced so many measures of information for belief functions, that by saying that every sensible measure has been defined, we would not be far from truth (for a review of this measures, see [23]). I shall consider two measures of information for beliefs: a measure of specificity and one of dissonance (for some very interesting considerations about some measures strictly related to these two, see [17]).

Definition 65 *The (measure of) specificity of a belief b on $\mathcal{U}$ is*

$$S(b) = -\sum_{\emptyset \neq s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b}(s) \log_2 \frac{|s|}{|V_{\mathcal{U}}|}.$$

The unit of this measure can be interpreted as the total specificity of the belief about a single elementary proposition (i.e., a proposition represented by a propositional symbol): if $b \in B_{\{a\}}$, then $S(b) = 1 - \left(p^{b}(a) - b(a)\right)$ (that is, $S(b) = 1 \Leftrightarrow b(a) = p^{b}(a)$). Consequently, the maximal possible specificity depends on the number of propositional symbols considered: for every belief b on $\mathcal{U}$, $S(b) \in [0, |\mathcal{U}|]$. Obviously, a belief on $\mathcal{U}$ has the maximal specificity $|\mathcal{U}|$ if and only if it is an epistemic probability (this is another characterization of this particular kind of beliefs), and has the minimal specificity 0 if and only if it is the vacuous belief. For instance, for $a, b \in \mathcal{U}$, $b_{\mathcal{U}}^{\langle a,\gamma\rangle} = \gamma$, $b_{\mathcal{U}}^{\langle (a \wedge b),\gamma\rangle} = 2\gamma$, and $b_{\mathcal{U}}^{\langle (a \leftrightarrow b),\gamma\rangle} = \gamma$. Consider two sets of propositional symbols $\mathcal{U} \subset \mathcal{V}$ and two beliefs $b_{\mathcal{U}}$ and $b_{\mathcal{V}}$ on $\mathcal{U}$ and $\mathcal{V}$, respectively: it is easy to show that if $b_{\mathcal{V}}$ is an extension of $b_{\mathcal{U}}$ to $\mathcal{V}$, then $S(b_{\mathcal{V}}) \geq S(b_{\mathcal{U}})$ (an extension can describe a more specific belief), and the equality holds if and only if the extension is vacuous ($S(b_{\mathcal{U}} \uparrow_{\mathcal{V}}) = S(b_{\mathcal{U}})$: this is a necessary requirement, since the difference between a belief and one of its vacuous extensions is purely formal). If $D = \{\langle \varphi_1, \alpha_1\rangle, \ldots, \langle \varphi_n, \alpha_n\rangle\}$ is a support distribution on $\mathcal{U}$ such that $\varphi_i \not\approx \perp$ for every $i \in \{1, \ldots, n\}$, then $S\left(b_{\mathcal{U}}^{D}\right) = -\sum_{i=1}^{n} \alpha_i \log_2 \frac{|s_{\mathcal{U}}(\varphi_i)|}{|V_{\mathcal{U}}|}$, where $\frac{|s_{\mathcal{U}}(\varphi_i)|}{|V_{\mathcal{U}}|}$ is the proportion of the valuations of $\mathcal{U}$ which are models of φ_i: the fewer models the supported formulas have, the more specific is the belief.

Proposition 66 *Given two beliefs b_1 and b_2 on $\mathcal{U}$ and an acceptable JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$,*

$$S\left(b_1 \otimes b_2\right) \geq \frac{\max\left(S\left(b_1\right), S\left(b_2\right)\right) - c\left(b_1 \otimes b_2\right)|\mathcal{U}|}{1 - c\left(b_1 \otimes b_2\right)}.$$

Proof.

$$\begin{aligned}
S\left(b_1 \otimes b_2\right) &= -\sum_{\emptyset \neq s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1 \otimes b_2}(s) \log_2 \tfrac{|s|}{|V_{\mathcal{U}}|} = \\
&= -\sum_{\emptyset \neq s \subset V_{\mathcal{U}}} \tfrac{1}{1-c(b_1 \otimes b_2)} \sum_{s' \cap s'' = s} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \log_2 \tfrac{|s|}{|V_{\mathcal{U}}|} = \\
&= -\tfrac{1}{1-c(b_1 \otimes b_2)} \sum_{s' \cap s'' \neq \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \log_2 \tfrac{|s|}{|V_{\mathcal{U}}|} \geq \\
&\geq -\tfrac{1}{1-c(b_1 \otimes b_2)} \sum_{\emptyset \neq s' \subset V_{\mathcal{U}}} \left(\sum_{s'' \subset V_{\mathcal{U}}} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \log_2 \tfrac{|s'|}{|V_{\mathcal{U}}|} - \right. \\
&\qquad \left. - \sum_{s' \cap s'' = \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \log_2 \tfrac{|s'|}{|V_{\mathcal{U}}|} \right) \geq \\
&\geq \tfrac{1}{1-c(b_1 \otimes b_2)} \left(-\sum_{\emptyset \neq s' \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b_1}\left(s'\right) \log_2 \tfrac{|s'|}{|V_{\mathcal{U}}|} + \sum_{s' \cap s'' = \emptyset} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}\left(s', s''\right) \log_2 \tfrac{1}{|V_{\mathcal{U}}|} \right) = \\
&= \tfrac{S(b_1) - c(b_1 \otimes b_2)|\mathcal{U}|}{1-c(b_1 \otimes b_2)}.
\end{aligned}$$

Analogously we obtain $S\left(b_1 \otimes b_2\right) \geq \frac{S(b_2) - c(b_1 \otimes b_2)|\mathcal{U}|}{1-c(b_1 \otimes b_2)}$, and thus the desired result. ■

This lower bound for $S\left(b_1 \otimes b_2\right)$ is often attained: therefore, when we combine two beliefs, because of conflict we can lose specificity (this is another indication of the unfavorableness of the conflict). Consider for instance the situation in which $\mathcal{U} = \{a, b\}$, $b_1 = b_{\mathcal{U}}^{\langle (a \vee b), 1 \rangle}$ and $b_2 = b_{\mathcal{U}}^{\left\{ \left\langle (a \vee b), \frac{1}{2} \right\rangle, \left\langle (\neg a \vee \neg b), \frac{1}{2} \right\rangle \right\}}$: $S\left(b_1\right) = \log_2 \frac{4}{3}$ and $S\left(b_2\right) = 1 + \frac{1}{2} \log_2 \frac{4}{3}$; since b_1 is a certainty, the JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ is unique ($c\left(b_1 \otimes b_2\right) = \frac{1}{2}$ and $b_1 \otimes b_2 = b_1$) and $\frac{\max(S(b_1), S(b_2)) - c(b_1 \otimes b_2)|\mathcal{U}|}{1-c(b_1 \otimes b_2)} = \frac{1 + \frac{1}{2}\log_2 \frac{4}{3} - \frac{1}{2}2}{1 - \frac{1}{2}} = \log_2 \frac{4}{3} = S\left(b_1\right) = S\left(b_1 \otimes b_2\right)$ (the lower bound of Proposition 66 is attained). If we consider the relative specificity $\frac{S(b)}{|\mathcal{U}|}$ instead of the specificity $S(b)$, the inequality of Proposition 66 becomes

$$\frac{S\left(b_1 \otimes b_2\right)}{|\mathcal{U}|} \geq \frac{\max\left(\frac{S(b_1)}{|\mathcal{U}|}, \frac{S(b_2)}{|\mathcal{U}|}\right) - c\left(b_1 \otimes b_2\right)}{1 - c\left(b_1 \otimes b_2\right)}:$$

the analogy with the first inequality of Theorem 58 is remarkable. From Proposition 66 it follows in particular that if $c\left(b_1 \otimes b_2\right) = 0$, then $S\left(b_1 \otimes b_2\right) \geq S\left(b_1\right)$ (and analogously for b_2): if there is no conflict, then the specificity is maintained. Another interesting special case is the one in which $S\left(b_1\right) = |\mathcal{U}|$ or $S\left(b_2\right) = |\mathcal{U}|$ (i.e., at least one of the two beliefs is an epistemic probability): $S\left(b_1 \otimes b_2\right) = |\mathcal{U}|$; that is, the maximal specificity cannot be lost (in other words: the combination of an epistemic probability with another belief is still an epistemic probability). It can be easily proved that, if $b_{\mathcal{U}}$ and $b_{\mathcal{V}}$ are two beliefs on

$\mathcal{U}$ and $\mathcal{V}$, respectively, and $\mathcal{U} \cap \mathcal{V} = \emptyset$, then for every acceptable JSSD $\underline{m}_{\mathcal{U}\cup\mathcal{V}}^{(b_\mathcal{U}\uparrow_{\mathcal{U}\cup\mathcal{V}})\otimes(b_\mathcal{V}\uparrow_{\mathcal{U}\cup\mathcal{V}})}$: $S\left((b_\mathcal{U}\uparrow_{\mathcal{U}\cup\mathcal{V}}) \otimes (b_\mathcal{V}\uparrow_{\mathcal{U}\cup\mathcal{V}})\right) = S(b_\mathcal{U}) + S(b_\mathcal{V})$ (thus, in a certain sense, the specificity about different topics is additive). Of course, the specificity of the weighted average of some beliefs is the weighted average of their specificities (with the same weights); in particular for the discounting (since the vacuous belief has specificity 0): $S(\gamma b) = \gamma S(b)$.

I shall now consider the second measure of information: the measure of dissonance.

Definition 67 *The (measure of) dissonance of a belief b on $\mathcal{U}$ is*

$$D(b) = - \sum_{s \subset V_\mathcal{U}:\ p^b(\varphi_s) \neq 0} m_\mathcal{U}^b(s) \log_2 p^b(\varphi_s),$$

where, for every $s \subset V_\mathcal{U}$, φ_s is a formula on $\mathcal{U}$ such that $s_\mathcal{U}(\varphi_s) = s$.

The unit of this measure can be interpreted as the total dissonance of the belief about a single elementary proposition: if $b \in B_{\{a\}}$, then $D(b) = 1 \Leftrightarrow b(a) = b(\neg a) = \frac{1}{2}$, and $D(b) = 0 \Leftrightarrow (b(a) = 0 \text{ or } b(\neg a) = 0)$. Consequently, as it is the case for the specificity, the maximal possible dissonance depends on the number of propositional symbols considered: for every belief b on $\mathcal{U}$, $D(b) \in [0, |\mathcal{U}|]$. But in contrast with the situation of the specificity (where the minimum is reached only by the vacuous belief, and the maximum by every epistemic probability), a belief on $\mathcal{U}$ has the maximal dissonance $|\mathcal{U}|$ if and only if it is the epistemic probability which assigns the same belief to every elementary event (i.e., from the semantical standpoint, a singleton of $V_\mathcal{U}$); that is, the so-called "uniform Bayesian prior". On the other hand, every consonant belief on $\mathcal{U}$ has the minimal dissonance 0 (that is, "consonant" implies "not dissonant"), but the minimal dissonance is reached also by not consonant beliefs (that is, "dissonant" is not the contrary of "consonant"): for instance, for $a, b \in \mathcal{U}$, $b_\mathcal{U}^{\{\langle a,\gamma\rangle,\langle b,\delta\rangle\}} = 0$. It is easy to show that $D(b) = 0 \Leftrightarrow \left(b(\varphi) > 0 \Rightarrow p^b(\varphi) = 1\right)$: we have seen in Section 1.9 that the right-hand side is necessary (but obviously not sufficient) for b to be consonant; another interesting characterization of the beliefs with minimal dissonance is the following: $D(b) = 0 \Leftrightarrow c(b \oplus b) = 0$. Thus the minimal dissonance does not characterize the consonant beliefs as the maximal specificity does for the epistemic probabilities; for these, the dissonance reduces to the well-known Shannon's entropy. Consider now two sets of propositional symbols $\mathcal{U} \subset \mathcal{V}$ and two beliefs $b_\mathcal{U}$ and $b_\mathcal{V}$ on $\mathcal{U}$ and $\mathcal{V}$, respectively: it is easy to show that if $b_\mathcal{V}$ is an extension of $b_\mathcal{U}$ to $\mathcal{V}$, then $D(b_\mathcal{V}) \geq D(b_\mathcal{U})$ (an extension can carry more dissonance); and if this extension is vacuous, then $D(b_\mathcal{U}\uparrow_\mathcal{V}) = D(b_\mathcal{U})$ (as noted for the specificity, this is a necessary requirement). In contrast with the case of specificity, the equality $D(b_\mathcal{V}) = D(b_\mathcal{U})$ does not characterize the vacuous extension; for instance, if $a \in \mathcal{U} \cap \mathcal{V}$ and $b \in \mathcal{V} \setminus \mathcal{U}$, then $b_\mathcal{V}^{\left\{\left\langle a,\frac{1}{2}\right\rangle,\left\langle (a\wedge b),\frac{1}{2}\right\rangle\right\}}$ is an extension (not vacuous) of $b_\mathcal{U}^{\{\langle a,1\rangle\}}$, but both beliefs have no dissonance (since they are consonant).

We could obtain a lower bound for the dissonance of a combination of beliefs, like the one for the specificity stated in Proposition 66; but the lower bound for the dissonance would be much more complicated and thus less interesting than the one for the specificity. Therefore I shall only prove that if there is no conflict, then the dissonance is maintained.

Proposition 68 *Given two beliefs b_1 and b_2 on $\mathcal{U}$ and a JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ such that $c(b_1 \otimes b_2) = 0$,*

$$D(b_1 \otimes b_2) \geq \max(D(b_1), D(b_2)).$$

Proof. From Theorem 58 we get $p^{b_1 \otimes b_2}(\varphi) \leq \min\left(p^{b_1}(\varphi), p^{b_2}(\varphi)\right) \leq p^{b_1}(\varphi)$; therefore, since (for every belief b on $\mathcal{U}$) $p^b(\varphi_s) = 0 \Rightarrow m_{\mathcal{U}}^b(s) = 0$,

$$\begin{aligned} D(b_1 \otimes b_2) &= -\sum_{s \subset V_{\mathcal{U}}:\ p^{b_1 \otimes b_2}(\varphi_s) \neq 0} m_{\mathcal{U}}^{b_1 \otimes b_2}(s) \log_2 p^{b_1 \otimes b_2}(\varphi_s) = \\ &= -\sum_{s \subset V_{\mathcal{U}}:\ p^{b_1 \otimes b_2}(\varphi_s) \neq 0} \sum_{s' \cap s'' = s} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s', s'') \log_2 p^{b_1 \otimes b_2}(\varphi_s) = \\ &= -\sum_{s', s'' \subset V_{\mathcal{U}}:\ p^{b_1 \otimes b_2}((\varphi_{s'} \wedge \varphi_{s''})) \neq 0} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s', s'') \log_2 p^{b_1 \otimes b_2}((\varphi_{s'} \wedge \varphi_{s''})) \geq \\ &\geq -\sum_{s', s'' \subset V_{\mathcal{U}}:\ p^{b_1 \otimes b_2}(\varphi_{s'}) \neq 0} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s', s'') \log_2 p^{b_1 \otimes b_2}(\varphi_{s'}) \geq \\ &\geq -\sum_{s' \subset V_{\mathcal{U}}:\ p^{b_1}(\varphi_{s'}) \neq 0} \sum_{s'' \subset V_{\mathcal{U}}} \underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}(s', s'') \log_2 p^{b_1}(\varphi_{s'}) = \\ &= -\sum_{s' \subset V_{\mathcal{U}}:\ p^{b_1}(\varphi_{s'}) \neq 0} m_{\mathcal{U}}^{b_1}(s') \log_2 p^{b_1}(\varphi_{s'}) = D(b_1). \end{aligned}$$

Analogously we obtain $D(b_1 \otimes b_2) \geq D(b_2)$, and thus the desired result. ■

Since Dempster's rule has a simple mathematical form, for the special case with $c(b_1 \oplus b_2) = 0$ we have that $D(b_1 \oplus b_2) \geq D(b_1) + D(b_2)$: the proof is analogous to the one of Proposition 68 (using the result of Theorem 60: $p^{b_1 \oplus b_2}(\varphi) \leq p^{b_1}(\varphi)\, p^{b_2}(\varphi)$). But in practice the special cases without conflict are not so interesting, because two dissonant beliefs are in general conflicting. The only situation in which both beliefs can be highly dissonant and not conflicting is the one in which the two beliefs are about two different topics: as can be easily shown, if $b_{\mathcal{U}}$ and $b_{\mathcal{V}}$ are two beliefs on $\mathcal{U}$ and $\mathcal{V}$, respectively, and $\mathcal{U} \cap \mathcal{V} = \emptyset$, then the dissonances are additive with respect to Dempster's rule: $D((b_{\mathcal{U}} \uparrow_{\mathcal{U} \cup \mathcal{V}}) \oplus (b_{\mathcal{V}} \uparrow_{\mathcal{U} \cup \mathcal{V}})) = D(b_{\mathcal{U}}) + D(b_{\mathcal{V}})$ (but this does not hold for the general combination $\otimes$). If there is conflict, then the dissonance can be lost; consider for instance (for $a, b \in \mathcal{U}$) the two beliefs $b_1 = b_{\mathcal{U}}^{\left\{\left\langle a, \frac{1}{2}\right\rangle, \left\langle \neg a, \frac{1}{2}\right\rangle\right\}}$ and $b_2 = b_{\mathcal{U}}^{\left\{\left\langle (a \wedge b), \frac{1}{2}\right\rangle, \left\langle (a \wedge \neg b), \frac{1}{2}\right\rangle\right\}}$: $D(b_1) = D(b_2) = 1$, but we can easily define an acceptable JSSD $\underline{m}_{\mathcal{U}}^{b_1 \otimes b_2}$ such that $D(b_1 \otimes b_2) = 0$. It is interesting to notice that in the case of two opposite beliefs, the combination with minimal conflict creates in general a high dissonance, whereas the combination $\odot$ keeps the dissonance low: for instance, if $\varphi \not\approx \top$, $D\left(b_{\mathcal{U}}^{\left\langle \varphi, \frac{1}{2}\right\rangle} \odot b_{\mathcal{U}}^{\left\langle \neg\varphi, \frac{1}{2}\right\rangle}\right) = 0$, whereas $D\left(b_{\mathcal{U}}^{\langle\varphi,\frac{1}{2}\rangle}\oplus b_{\mathcal{U}}^{\langle\neg\varphi,\frac{1}{2}\rangle}\right)=\frac{2}{3}\log_2\frac{3}{2}$, and for the combination with minimal conflict (which in this case is unique) $D\left(b_{\mathcal{U}}^{\langle\varphi,\frac{1}{2}\rangle}\otimes b_{\mathcal{U}}^{\langle\neg\varphi,\frac{1}{2}\rangle}\right)=1$. From Jensen's inequality, we obtain easily that the discounting of a belief reduces strongly its dissonance: $D(\gamma b)\leq\gamma^2 D(b)$, where the equality holds only if $D(b)=0$ or $\gamma\in\{0,1\}$.

Theorem 69 states that the specificity of a belief is always at least as large as than its dissonance (this result was first proven by Ramer in [25]). Therefore, for every belief b on $\mathcal{U}$, $(S(b)-D(b))\in[0,|\mathcal{U}|]$: the definition of $S-D$ as a new measure of information is very tempting, but I cannot see a clear interpretation of this measure. Besides, the measure $S-D$ presents a problem: if we have two sets of propositional symbols $\mathcal{U}\subset\mathcal{V}$ and two beliefs $b_{\mathcal{U}}$ and $b_{\mathcal{V}}$ on $\mathcal{U}$ and $\mathcal{V}$, respectively, such that $b_{\mathcal{V}}$ is an extension of $b_{\mathcal{U}}$ to $\mathcal{V}$, then in general the inequality $S(b_{\mathcal{V}})-D(b_{\mathcal{V}})\geq S(b_{\mathcal{U}})-D(b_{\mathcal{U}})$ is not valid (this is hardly acceptable, since an extension carries more information); whereas, at least, the equality $S(b_{\mathcal{U}}{\uparrow}_{\mathcal{V}})-D(b_{\mathcal{U}}{\uparrow}_{\mathcal{V}})=S(b_{\mathcal{U}})-D(b_{\mathcal{U}})$ holds (this is certainly a strictly necessary requirement). However, since the specificity is a positive aspect of a belief and the dissonance a negative one, the measure $S-D$ could be very useful from a practical standpoint. Obviously, for a belief b on $\mathcal{U}$, the measure $S(b)-D(b)$ reaches its maximum $|\mathcal{U}|$ if and only if b is the certainty of φ and $|s_{\mathcal{U}}(\varphi)|=1$ (i.e., φ has exactly one model); that is, $S-D$ is maximal exactly for those beliefs which are simultaneously consonant beliefs and epistemic probabilities. Also the beliefs for which the measure $S-D$ reaches its minimum 0 are very interesting: from the proof of Theorem 69 it is easy to see that, for $b\in B_{\mathcal{U}}$, $S(b)-D(b)$ is minimal if and only if the elements of $\left\{s\subset V_{\mathcal{U}}\mid m_{\mathcal{U}}^{b}(s)\neq 0\right\}$ build a partition of $V_{\mathcal{U}}$ and for these sets $m_{\mathcal{U}}^{b}(s)=\frac{|s|}{|V_{\mathcal{U}}|}$. That is, $S-D$ reaches its minimum for those beliefs which present an extreme symmetry: in particular, the only consonant belief that minimize $S-D$ is the vacuous belief, and the only epistemic probability that minimize $S-D$ is the uniform Bayesian prior.

Theorem 69 *Given a belief b on $\mathcal{U}$,*

$$S(b)\geq D(b).$$

Proof. Since $m_{\mathcal{U}}^{b}(s) \neq 0 \Rightarrow p^{b}(\varphi_s) \neq 0$,

$$\begin{aligned}
S(b) - D(b) &= -\sum_{\emptyset \neq s \subset V_{\mathcal{U}}} m_{\mathcal{U}}^{b}(s) \log_2 \tfrac{|s|}{|V_{\mathcal{U}}|} - \sum_{s \subset V_{\mathcal{U}}:\ p^{b}(\varphi_s) \neq 0} m_{\mathcal{U}}^{b}(s) \log_2 p^{b}(\varphi_s) = \\
&= -\sum_{s \subset V_{\mathcal{U}}:\ p^{b}(\varphi_s) \neq 0} m_{\mathcal{U}}^{b}(s) \log_2 \tfrac{|s|}{p^{b}(\varphi_s)|V_{\mathcal{U}}|} \geq \\
&\geq -\log_2 \left(\sum_{s \subset V_{\mathcal{U}}:\ p^{b}(\varphi_s) \neq 0} |s| \tfrac{m_{\mathcal{U}}^{b}(s)}{p^{b}(\varphi_s)|V_{\mathcal{U}}|} \right) = \\
&= -\log_2 \left(\sum_{s \subset V_{\mathcal{U}}:\ m_{\mathcal{U}}^{b}(s) \neq 0} \sum_{v \in s} \tfrac{m_{\mathcal{U}}^{b}(s)}{p^{b}(\varphi_s)|V_{\mathcal{U}}|} \right) = \\
&= -\log_2 \left(\sum_{v \in V_{\mathcal{U}}} \sum_{s \ni v:\ m_{\mathcal{U}}^{b}(s) \neq 0} \tfrac{m_{\mathcal{U}}^{b}(s)}{p^{b}(\varphi_s)|V_{\mathcal{U}}|} \right) \geq \\
&\geq -\log_2 \left(\sum_{v \in V_{\mathcal{U}}} \sum_{s \ni v:\ m_{\mathcal{U}}^{b}(s) \neq 0} \frac{m_{\mathcal{U}}^{b}(s)}{\left(\sum_{s' \ni v} m_{\mathcal{U}}^{b}(s') \right)|V_{\mathcal{U}}|} \right) = \\
&= -\log_2 \left(\sum_{v \in V_{\mathcal{U}}:\ \left(\exists s \ni v:\ m_{\mathcal{U}}^{b}(s) \neq 0 \right)} \tfrac{1}{|V_{\mathcal{U}}|} \right) \geq -\log_2 \left(\sum_{v \in V_{\mathcal{U}}} \tfrac{1}{|V_{\mathcal{U}}|} \right) = \\
&= -\log_2 1 = 0;
\end{aligned}$$

where the first inequality is a special case of Jensen's one,
and the second one is based on $v \in s \Rightarrow p^{b}(\varphi_s) \geq \sum_{s' \ni v} m_{\mathcal{U}}^{b}(s')$. ■

2.6 Conditioning

The conditioning is a particular case of combination: given a belief b on $\mathcal{U}$ and a formula $\varphi \in \mathcal{L}_{\mathcal{U}}$, the conditioning of b on φ is the belief $b \otimes b_{\mathcal{U}}^{\langle \varphi, 1 \rangle}$, which is defined if $c\left(b \otimes b_{\mathcal{U}}^{\langle \varphi, 1 \rangle}\right) = b(\neg\varphi) < 1 \Leftrightarrow p^{b}(\varphi) > 0$ (i.e., the conditioning of b on φ is possible, if φ is considered possible by b). Remember that in the case of the combination of a belief and a certainty, the JSSD is unique: $b \otimes b_{\mathcal{U}}^{\langle \varphi, 1 \rangle}$ is thus well-defined. On one hand, thanks to the uniqueness of the JSSD, the conditioning does not need the notion of the independence of the sources of information, and thus is the clearest kind of combination: therefore it is often considered as the basis of the dynamic aspect of the belief function theory (in fact it is considered as an axiom by both axiomatic derivations of Dempster's rule of combination cited in Section 2.2). On the other hand, since it is the only kind of combination which has a direct counterpart in the Bayesian probability theory, the conditioning is the aspect of the Dempster-Shafer theory most subject to criticisms.

Given a belief b on $\mathcal{U}$ and a formula $\varphi \in \mathcal{L}_{\mathcal{U}}$ such that $p^b(\varphi) > 0$,

$$\left(b \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi) = \frac{b\left((\varphi \rightarrow \psi)\right) - b(\neg\varphi)}{p^b(\varphi)} \text{ and } p^{b \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}}(\psi) = \frac{p^{b \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}}\left((\varphi \wedge \psi)\right)}{p^{b \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}}(\varphi)}.$$

These expressions have an appealing clarity, which supports the use of the conditioning as the basis of the dynamic aspect of the Dempster-Shafer theory. The first one tells us that our belief in ψ after the acquisition of the certainty of φ is the belief we had in "If φ, then ψ" without the part of this one which was given directly to $\neg\varphi$ (notice that $(\varphi \rightarrow \psi) \sim (\neg\varphi \vee \psi) \sim (\neg\varphi \vee (\varphi \wedge \psi)))$; the whole normalized through the plausibility we gave to φ. In particular if we did not impugn φ at all (i.e., $p^b(\varphi) = 1$), then $\left(b \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi) = b\left((\varphi \rightarrow \psi)\right)$: the belief in the logical implication "If φ, then ψ" is the belief in ψ after the acquisition of the certainty of φ; this interpretation of the logical implication is extremely interesting: from the logical standpoint this is surely one of the most important advantages of the belief functions theory with respect to the Bayesian probability theory (in fact it allows to express default logic as a special case of the Dempster-Shafer theory). The second expression tells us that the plausibilities behave like probabilities in the case of conditioning; in particular from this expression it follows directly that for epistemic probabilities the conditioning of beliefs corresponds to the Bayesian one (since for epistemic probabilities $b = p^b$ holds).

If $D = \{\langle\varphi_1, \alpha_1\rangle, \ldots, \langle\varphi_n, \alpha_n\rangle\}$ is a complete support distribution and φ a formula, both on $\mathcal{U}$, then the conditioning of $b_{\mathcal{U}}^D$ on φ (if it is possible) is the belief $b_{\mathcal{U}}^{D'}$, where $D' = \{\langle(\varphi_1 \wedge \varphi), \alpha_1\rangle, \ldots, \langle(\varphi_n \wedge \varphi), \alpha_n\rangle\}$: if $\varphi_i \not\rightsquigarrow \neg\varphi$, then the support which was assigned to φ_i is transferred to $(\varphi_i \wedge \varphi)$, and then the whole is renormalized. If φ_i has more than one model (i.e., $|s_{\mathcal{U}}(\varphi_i)| > 1$), then α_i is the support which we assigned to φ_i without having enough knowledge to assign it to formulas ψ such that $\psi \rightsquigarrow \varphi_i$ (i.e., $s_{\mathcal{U}}(\psi) \subset s_{\mathcal{U}}(\varphi_i)$). With the conditioning on φ, in the case in which $\varphi_i \not\rightsquigarrow \varphi$ and $\varphi_i \not\rightsquigarrow \neg\varphi$ (i.e., $s_{\mathcal{U}}(\varphi_i) \cap s_{\mathcal{U}}(\varphi) \neq \emptyset$ and $s_{\mathcal{U}}(\varphi_i) \setminus s_{\mathcal{U}}(\varphi) \neq \emptyset$), α_i is transferred completely to $(\varphi_i \wedge \varphi)$; without considering that if we have had more knowledge, perhaps we would have assigned this support to a formula ψ such that $\psi \rightsquigarrow (\varphi_i \wedge \neg\varphi)$ (i.e., $s_{\mathcal{U}}(\psi) \subset s_{\mathcal{U}}(\varphi_i) \setminus s_{\mathcal{U}}(\varphi)$). In this sense the conditioning of beliefs is very incautious; we shall see that this incautiousness is at the same time a weakness and a strength of the Dempster-Shafer theory.

The incautiousness of the conditioning of beliefs is a problem when dealing with incomplete statistical information. Consider the following example: we are interested in an individual x, and in particular we are concerned with the question if x presents the features A and B or if it does not; we can describe our belief about this problem by considering the propositional symbols a and b (i.e., $\mathcal{U} = \{a, b\}$), where a represents the proposition "x presents the feature A" and b the proposition "x presents the feature B". Assume that the only available source of knowledge is a statistic which tells us that in a collection X of individuals like x, exactly one half presents the feature A (and the other half does not), and of the individuals which present the feature A, a proportion of

$\frac{9}{10}$ presents also the feature B (and the remaining $\frac{1}{10}$ does not); thus the statistic tells us nothing about the proportion of individuals presenting B among those which do not present A (in this sense the statistic is incomplete). The belief we get from this statistic is $b_{\mathcal{U}}^{\left\{\left\langle(a\wedge b),\frac{9}{20}\right\rangle,\left\langle(a\wedge\neg b),\frac{1}{20}\right\rangle,\left\langle\neg a,\frac{1}{2}\right\rangle\right\}}$; if now we obtain (from another source of information) the certainty of $\neg b$ (that is, we are sure that x does not present the feature B), the belief after the conditioning becomes $b_{\mathcal{U}}^{\left\{\left\langle(a\wedge\neg b),\frac{1}{11}\right\rangle,\left\langle(\neg a\wedge\neg b),\frac{10}{11}\right\rangle\right\}}$, in particular the belief in $\neg a$ becomes $\frac{10}{11}$. This is debatable, since if the statistic had told us also the proportion $\alpha \in \left[0, \frac{1}{2}\right]$ of the elements of X presenting B but not presenting A, the initial belief would have been the epistemic probability $b_{\mathcal{U}}^{\left\{\left\langle(a\wedge b),\frac{9}{20}\right\rangle,\left\langle(a\wedge\neg b),\frac{1}{20}\right\rangle,\langle(\neg a\wedge b),\alpha\rangle,\left\langle(\neg a\wedge\neg b),\frac{1}{2}-\alpha\right\rangle\right\}}$ and the belief in $\neg a$ after the (Bayesian) conditioning would have been $\frac{\frac{1}{2}-\alpha}{\frac{11}{20}-\alpha}$, whose range is $\left[0, \frac{10}{11}\right]$. Therefore the belief in $\neg a$ obtained without knowing the proportion α is the maximal possible one among those which would have been obtained if this proportion had been known: this is due to the incautiousness of the conditioning of beliefs.

It is very important to notice that in the above situation the belief on $\mathcal{U}$ represents a lower probability, in the sense that, for every formula ψ on $\mathcal{U}$, $b(\psi) = \min_{b' \in P_{\mathcal{U}}^b} b'(\psi)$, where $P_{\mathcal{U}}^b = \left\{ b_{\mathcal{U}}^{\left\{\left\langle(a\wedge b),\frac{9}{20}\right\rangle,\left\langle(a\wedge\neg b),\frac{1}{20}\right\rangle,\langle(\neg a\wedge b),\alpha\rangle,\left\langle(\neg a\wedge\neg b),\frac{1}{2}-\alpha\right\rangle\right\}} \mid \alpha \in \left[0, \frac{1}{2}\right] \right\}$ (every element of $P_{\mathcal{U}}^b$ is an epistemic probability on $\mathcal{U}$). Every belief can formally be seen as a lower probability, and the associated plausibility as the respective upper probability (lower and upper probabilities are the object of the original work of Dempster [4], see also [19]): given a belief b on $\mathcal{U}$, let $P_{\mathcal{U}}^b \subset B_{\mathcal{U}}$ be the set of the epistemic probabilities b' on $\mathcal{U}$ such that $b' \geq b$ (notice that $b' \in P_{\mathcal{U}}^b \Rightarrow b \leq b' \leq p^b$); it is easy to show that, for every formula ψ on $\mathcal{U}$, $b(\psi) = \min_{b' \in P_{\mathcal{U}}^b} b'(\psi)$ and $p^b(\psi) = \max_{b' \in P_{\mathcal{U}}^b} b'(\psi)$. If we would like to avoid the problems of the kind of the example above, then obviously we should define the belief on ψ after the acquisition of the certainty of φ as the lower conditional probability $\min_{b' \in P_{\mathcal{U}}^b} \left(b' \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi)$, and the plausibility as the upper conditional probability $\max_{b' \in P_{\mathcal{U}}^b} \left(b' \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi)$ (of course these are defined only if $p^{b'}(\varphi) > 0$ for every $b' \in P_{\mathcal{U}}^b$; i.e., if $b(\varphi) > 0$). It can be shown (see for instance [12]) that (if $b(\varphi) > 0$) these two functions are a belief on $\mathcal{U}$ and the associated plausibility, and that the following equalities hold:

$$\begin{aligned}
\min_{b' \in P_{\mathcal{U}}^b} \left(b' \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi) &= \frac{b((\varphi \wedge \psi))}{b((\varphi \wedge \psi)) + p^b((\varphi \wedge \neg\psi))} \text{ and} \\
\max_{b' \in P_{\mathcal{U}}^b} \left(b' \otimes b_{\mathcal{U}}^{\langle\varphi,1\rangle}\right)(\psi) &= \frac{p^b((\varphi \wedge \psi))}{p^b((\varphi \wedge \psi)) + b((\varphi \wedge \neg\psi))}.
\end{aligned}$$

These expressions are not so clearly interpretable as the two about the conditioning of beliefs are; but it is easy to verify that if b is an epistemic probability, they express the

Bayesian conditioning (this is trivial if we look at the left side of the equations). As the conditioning of a belief is very incautious, this way of changing a belief after the acquisition of a certainty is very cautious: in the above example the final belief in $\neg a$ would have been 0; in fact, from the Bayesian standpoint, the belief 0 is the best we can affirm with certainty, but to require such a certainty when we are dealing with inductive reasoning is surely debatable. It is easy to show that the solution obtained with the lower conditional probability is always more cautious than the one obtained with the conditioning of beliefs: for every $\varphi, \psi \in \mathcal{L}_{\mathcal{U}}$ such that $b(\varphi) > 0$,

$$\min_{b' \in P^b_{\mathcal{U}}} \left(b' \otimes b^{\langle \varphi, 1 \rangle}_{\mathcal{U}} \right)(\psi) \leq \left(b \otimes b^{\langle \varphi, 1 \rangle}_{\mathcal{U}} \right)(\psi) \leq p^{b \otimes b^{\langle \varphi, 1 \rangle}_{\mathcal{U}}}(\psi) \leq \max_{b' \in P^b_{\mathcal{U}}} \left(b' \otimes b^{\langle \varphi, 1 \rangle}_{\mathcal{U}} \right)(\psi).$$

Since the lower conditional probability approach is supported by the Bayesian probability theory, some authors (in particular Dubois and Prade, see for instance [10]) have suggested an extensive utilization of this approach in the belief functions theory, without considering that from the conceptual standpoint it does not match with the rest of the theory. Following Dubois and Prade, given a belief b and a formula φ on $\mathcal{U}$ such that $b(\varphi) > 0$, I shall call the belief $\tilde{b}$ on $\mathcal{U}$ defined through $\tilde{b}(\psi) = \min_{b' \in P^b_{\mathcal{U}}} \left(b' \otimes b^{\langle \varphi, 1 \rangle}_{\mathcal{U}} \right)(\psi)$ (for every formula ψ on $\mathcal{U}$) the focusing of b on φ. The problem of the focusing is that it does not match with the (Dempster's) combination rule (which is the core of the Dempster-Shafer theory); thus its use in the description of the evolution of a belief leads to absurdities (but this does not mean that the focusing cannot be useful for other purposes). First of all, there is no continuity: consider the combination of b with the simple support $b^{\langle \varphi, \alpha \rangle}_{\mathcal{U}}$, the limit of $b \oplus b^{\langle \varphi, \alpha \rangle}_{\mathcal{U}}$ when α tends to 1 is the conditioning of b on φ, not the focusing (and this holds also for the general combination rule, since the more we approach the certainty the less we have freedom in the choice of the JSSD); this problem could be solved (in a way not very convincing) by stressing the particularity of the certainties between the beliefs. But there are other problems; consider for instance the following situation (where $a, b \in \mathcal{U}$): from two independent sources of information (which are completely reliable) I get the beliefs $b^{\langle a, \alpha \rangle}_{\mathcal{U}}$ and $b^{\langle b, \beta \rangle}_{\mathcal{U}}$, respectively (so my belief will be $b^{\langle a, \alpha \rangle}_{\mathcal{U}} \oplus b^{\langle b, \beta \rangle}_{\mathcal{U}}$); afterwards the second source of information becomes more precise and tells me that b is certain. It seems obvious to me that the belief on a should be α (or at least it should not depend on β): in fact this is the solution if we use the conditioning (even if the independence was not assumed, i.e., this is the solution for any JSSD), whereas if we use the focusing, the belief in a is $\frac{\alpha\beta}{\alpha\beta + 1 - \alpha}$; notice that (if $\alpha < 1$) the limit of $\frac{\alpha\beta}{\alpha\beta + 1 - \alpha}$ when β tends to 0 ($\beta = 0$ is not allowed for the focusing) is 0: the less we knew about b the less we know about a. I think that such results would be a sufficient reason for rejecting the focusing rule, but there are also other problems; in particular the following one: the focusing of b on φ is defined only for $b(\varphi) > 0$ (i.e., we can accept only the certainties about propositions which we already support), and it cannot be extended to the case $b(\varphi) = 0$ without leading to discontinuities, as we can see from the previous example: if the result is continuous when

β tends to 0, then the case $\beta = 0$ gives the result 0 if $\alpha < 1$ and 1 if $\alpha = 1$ (differently from the previously stated discontinuity, this one is inside the focusing rule).

Consider the following situation: I believe with the same strength that it is raining in Yalta or that it is not, but if it is raining, then almost certainly the wind is blowing. The wind is not blowing in Yalta; what is the immediate conclusion that almost everyone would draw in this situation? Obviously, it is "very likely it is not raining in Yalta". This situation could be described with exactly the same model as the example of the incomplete statistical information: a represents the proposition "It is raining in Yalta" and b the proposition "The wind is blowing in Yalta". In this case too, the conditioning is incautious: I do not have any belief about the wind if it is not raining in Yalta; but in my opinion in this case the belief of $\frac{10}{11}$ in $\neg a$ after the conditioning on $\neg b$ is a good description of the human reasoning. This is what I meant when I said that the incautiousness of the conditioning is at the same time a weakness and a strength of the Dempster-Shafer theory: often the conclusions which seem senseless when carefully analyzed are the most spontaneous ones (for this reason the belief functions, which originated in the sphere of statistics, are studied and utilized almost exclusively in the sphere of artificial intelligence). Obviously, the correctness of these immediate conclusions is questionable. But an answer to the question about correctness cannot be given until a definition of "correct inductive inference" is available; and I think that such a definition is impossible, because it cannot be based on a requirement of infallibility (since the fallibility is what characterize the inductive inferences). In my opinion, "normal" human reasoning is correct, and thus I consider immediate conclusions like the one described above as being a form of "correct inductive inference" (but this is only my opinion, which is perhaps too optimistic).

Of course, we cannot pretend that exactly the same model gives us different results when we interpret it in different ways. If in the example of the incomplete statistical information we forget the statistical origin of our belief, the conclusion is sensible as it is in the example of the weather situation in Yalta; thus the conditioning is correct: if b is our belief and we get the certainty of φ, then the conditioning of b on φ is a good conclusion. The problem is not the statistical origin of our belief: also the belief about the weather in Yalta could have a statistical origin, and this would not change the sensibleness of the conclusion. The problem is that we accept the information which we obtain from the statistic as our belief; that is, we believe that the statistic is completely reliable: this means that we believe that there is an ontological probability that an individual like x presents the features A and B, respectively, and we believe that the collection X represents perfectly the individuals like x. From what I have said in Section 1.9, it is clear that in my opinion believing that a statistic is completely reliable is a nonsense (except if this complete reliability is an approximation of an extreme situation); therefore I think that every statistical knowledge should be discounted. In the example, if our belief in the reliability of the statistic was γ (i.e., the discount rate is $1 - \gamma$), then our final belief in $\neg a$ would be $\alpha = \frac{\frac{1}{2}\gamma}{1-\frac{9}{20}\gamma}$: thus the discounting has a very strong effect. For instance, if

$\gamma = \frac{4}{5} = 0.8$, then $\alpha = \frac{5}{8} = 0.625$; and even if $\gamma = \frac{19}{20} = 0.95$ (i.e., we believed the statistic to be extremely reliable), then $\alpha = \frac{190}{229} \cong 0.83$, which is a little smaller than the result $\alpha = \frac{5}{6}$ that we would obtain by choosing as our initial belief the epistemic probability which minimize the dissonance (i.e., Shannon's entropy) among the completions of the statistic. It is also important to stress that, to allow sensible results, a belief should be as complete as possible (i.e., the specificity should be as large as possible). Too often in the literature on belief functions only the vacuous belief and the epistemic probabilities are considered; but an unknown probability does not imply total ignorance (notice that the vacuous belief and the epistemic probabilities are the two extreme cases of the measure of specificity).

I shall consider now the well-known "three prisoners problem": it originated as a puzzle for the Bayesian probability theory (see [22]), and then it has been treated also as a puzzle for the Dempster-Shafer theory of evidence (see [6]). For a lively discussion among the supporters of the two theories, see the two special issues of the International Journal of Approximate Reasoning cited in Section 1.8: in the first one Pearl has considered this problem, and in the second one several authors have explained their interpretations; for a comparison of the results obtained using the conditioning and the focusing, respectively, see for instance [12].

Example 70 *The three prisoners problem for belief functions can be stated as follows: of the three prisoners Freeman, Grey and Hope, two are to be executed; and the judge randomly chooses which ones (this detail has been added in the formulation of the problem for belief functions, since in the Bayesian formulation the randomness is included in the uniform Bayesian prior). We consider the viewpoint of Freeman (who knows the situation that I have just described): he says to the jailer (who knows the decision of the judge), "Since either Grey or Hope is certainly going to be executed, you will give me no information about my own chances if you give me the name of one man, either Grey or Hope, who is going to be executed". Accepting this argument, the jailer truthfully replies "Grey will be executed". How does this information change the belief of Freeman about the men who are going to be executed? Using the propositional symbols f, g and h to represent the propositions "Freeman will be executed", "Grey will be executed" and "Hope will be executed", respectively, the belief b of Freeman on $\mathcal{U} = \{f, g, h\}$ must certainly satisfy the condition $b\left(\bigvee\left\{\bigwedge\{f, g, \neg h\}, \bigwedge\{f, \neg g, h\}, \bigwedge\{\neg f, g, h\}\right\}\right) = 1$ (i.e., he is sure that exactly two prisoners will be executed). Since Freeman knows that the judge chooses randomly, b should be symmetrical with respect to the three elementary propositions $\bigwedge\{f, g, \neg h\}$, $\bigwedge\{f, \neg g, h\}$ and $\bigwedge\{\neg f, g, h\}$. From this symmetry follows that if we combine b with the belief that Freeman gets from the jailer (the certainty of g), then the belief obtained is the same for f and h: $\left(b \otimes b_{\mathcal{U}}^{\langle g,1\rangle}\right)(f) = \left(b \otimes b_{\mathcal{U}}^{\langle g,1\rangle}\right)(h)$, and (unless b was the vacuous belief) both beliefs have increased. At first sight, this results seems sensible; in fact this is almost certainly the first answer that everyone would give. But, after an attentive analysis of the problem, this solution appears debatable, since something distinguishes Freeman from*

Hope: the second one, unlike the first one, could have been named by the jailer, and the fact that he was not named renders him suspect of being the one whose life will be spared. To introduce this difference between Freeman and Hope in the beliefs of the first one, we must consider what the jailer could have said: let j represent the proposition "The jailer names Grey" (thus $\neg j$ can be interpreted as "The jailer names Hope"); the belief of Freeman before the answer of the jailer (i.e., before the conditioning on j) should be on $\mathcal{U} \cup \{j\}$. The belief b can be extended to $\mathcal{U} \cup \{j\}$ through a conditional belief $b \mid \mathcal{U}$ on $\{j\}$, which should satisfy the conditions $b \mid \bigwedge \{f, g, \neg h\} = b_{\{j\}}^{\langle j,1\rangle}$ (i.e., if Freeman and Grey will be executed, then the jailer will name Grey) and $b \mid \bigwedge \{f, \neg g, h\} = b_{\{j\}}^{\langle \neg j,1\rangle}$ (i.e., if Freeman and Hope will be executed, then the jailer will name Hope), since it is assumed that Freeman knows that the jailer answers truthfully (I accept this assumption because by questioning it I shall change the classical problem, but please do not ask me how Freeman can be sure of this truthfulness). With this model, the final belief of Freeman about the men who are going to be executed is $b' = \left((b \otimes b \mid \mathcal{U}) \otimes b_{\mathcal{U}\cup\{j\}}^{\langle j,1\rangle}\right) \downarrow_{\mathcal{U}}$. In the literature on belief functions, the beliefs of Freeman b and $b \mid \mathcal{U}$ are always assumed to be the ones completely defined from $b(f) = b(g) = b(h) = \frac{1}{3}$ and $b \mid \bigwedge \{\neg f, g, h\} = b_{\{j\}}^{\emptyset}$; thus the final belief b' is the one such that $b'(f) = b'(h) = \frac{1}{2}$, which is subject to the above stated criticism. No author questions these choices: as I said before, "too often in the literature on belief functions only the vacuous belief and the epistemic probabilities are considered, but an unknown probability does not imply total ignorance". I think that the choice of the uniform Bayesian prior for b is acceptable (at least as an approximation), even if a little discounting would surely be better, considering also that Freeman has got information only by hearsay (he cannot be completely sure that the judge has chosen randomly). The real problem is the choice of the vacuous belief for $b \mid \bigwedge \{\neg f, g, h\}$: this means that Freeman gives absolutely no support to the possibility that the jailer would name Hope. The criticism of the first solution was precisely that Hope, unlike Freeman, could have been named by the jailer: obviously the fact that he was not named increases Freeman's belief in $\neg h$ only if Freeman gave some support to this possibility. In fact if we accept the choice of the uniform Bayesian prior for b and we set $b \mid \bigwedge \{\neg f, g, h\} = b_{\{j\}}^{\{\langle j,\alpha\rangle, \langle \neg j,\alpha\rangle\}}$ for an $\alpha \in \left(0, \frac{1}{2}\right]$ (i.e., Freeman gave the support α to $\neg j$, and symmetrically also to j), then the final belief b' satisfies $b'(f) = \frac{1-\alpha}{2-\alpha}$ and $b'(h) = \frac{1}{2-\alpha}$; thus $b'(h) > b'(f)$ and $\frac{b'(h)}{b'(f)} = \frac{1}{1-\alpha}$, whose range is $(1, 2]$ (the larger α is, the larger is $b'(h)$ with respect to $b'(f)$). I think that this solution solves the problem of the first one and clarifies the effect on the final belief of the support that was given to $\neg j$.

2.7 Bayes' Theorem

As I have said in Section 1.8, it is possible to formulate generalizations of Bayes' theorem for belief functions: for example the one proposed by Smets (see for instance [32] or

[30]). Thanks to the introduction of the notion of conditional belief (Section 2.2) and to the considerations about its utilization in relation with a metabelief (Section 2.4), we can easily obtain a generalization of Bayes' theorem in a way which is much clearer and allows fewer restrictive assumptions than the one used by Smets. Consider the following situation: we have two disjoint topics $\mathcal{U}$ and $\mathcal{V}$ (i.e., $\mathcal{U} \cap \mathcal{V} = \emptyset$), a prior belief b^{pri} on $\mathcal{V}$ (which could be vacuous, of course) and a conditional belief $b \mid \mathcal{V}$ on $\mathcal{U}$, then from a source of information we obtain a belief on $\mathcal{U}$ which we discount if the source is not completely reliable; let b be the discounted belief:

$$b^{post} = \left(\left(b^{pri} \otimes (b \mid \mathcal{V})\right) \otimes b\right){\downarrow_{\mathcal{V}}}$$

is the posterior belief on $\mathcal{V}$ (I have written b instead of $(b{\uparrow_{\mathcal{U}\cup\mathcal{V}}})$ to simplify the notation; also $b^{pri} \otimes (b \mid \mathcal{V})$ is the notation without vacuous extensions introduced in Section 2.2). Obviously this generalizes Bayes' theorem: if b^{pri} and (for every $\varphi \in \mathcal{L}_{\mathcal{V}}$ such that $s_{\mathcal{V}}(\varphi) = 1$) $b \mid \varphi$ are epistemic probabilities, and b is a certainty, then b^{post} is the epistemic probability which we would get from Bayes' theorem.

Smets' generalization considers only a particular situation, which I shall now study: I shall formulate it in terms of the inductive inference from some observations. Let $h_1, \ldots, h_n$ be some (scientific) hypotheses about a topic $\mathcal{U}$: every hypothesis h_i leads to a belief b_i on $\mathcal{U}$; then let b be the belief that we get from some observations (as before, I assume that b is already discounted). We want to study the effect of the observations on our belief about the correctness of the hypotheses, therefore let b^{pri} be our prior belief on $\mathcal{H} = \{h_1, \ldots, h_n\}$; to obtain the situation described above, we must assume that $\mathcal{U} \cap \mathcal{H} = \emptyset$ and define the conditional belief $b \mid \mathcal{H}$. Since b^{pri} can be seen as a metabelief about the reliability of the sources of information (i.e., the hypotheses) which give the beliefs $b_1, \ldots, b_n$, the construction of $b \mid \mathcal{H}$ has already been discussed in Section 2.4; but this construction allows a lot of freedom in the choice of $b \mid \mathcal{H}$. I shall study only a particular situation: I assume that the hypotheses are mutually excluding. This assumption means that

$$b^{pri}\left(\bigvee\left\{\bigwedge\{\varphi_1, \ldots, \varphi_n\} \mid \varphi_i \in \{h_i, \neg h_i\}\,,\ |\{i \mid \varphi_i \equiv h_i\}| \leq 1\right\}\right) = 1;$$

thus for instance $b^{pri}((h_1 \wedge h_2)) = 0$ and $b^{pri}((h_1 \vee h_2)) = b^{pri}((h_1 \wedge \neg h_2) \vee (\neg h_1 \wedge h_2))$: this is the mutual exclusion, which is assumed also in the Bayesian formulation and in Smets' generalization. But $b^{pri}(\bigwedge\{\neg h_1, \ldots, \neg h_n\})$ can be positive: we are not compelled to believe that the truth is among the hypotheses as we are in the Bayesian formulation and in Smets' generalization, in which however the same results can be obtained by adding a fictive hypothesis. With this assumption, the freedom in the choice of $b \mid \mathcal{H}$ is considerably reduced, in the sense that a lot of beliefs, like for instance $b \mid (h_1 \wedge h_2)$ will not be used in the combination $b^{pri} \otimes (b \mid \mathcal{H})$, and therefore their choice is insignificant. The only beliefs $b \mid \varphi$ which must be chosen are the ones such that φ is a disjunction of hypotheses: $\varphi = \bigvee\{h_i \mid i \in I\}$ for an $I \subset \{1, \ldots, n\}$ such that $|I| > 1$; I see only two possibilities for the choices of these beliefs without considering the particular situations: the vacuous belief or the disjunctive rule of combination. I consider only the second possibility, since it is a lot more interesting, and therefore I set $(b \mid \bigvee \{h_i \mid i \in I\}) = \biguplus_{i \in I} b_i$ (I think that this notation is clear, and it is sensible since the disjunctive rule is associative). From what I have said in Section 2.4, it follows that instead of the conditional belief we can use the hypothetical beliefs (in this case the name fits perfectly), and thus obtain

$$b^{post} = \left(\left(b^{pri} \oplus \bigoplus_{i=1}^{n} (h_i \rightharpoonup b_i) \right) \otimes b \right) \downarrow_{\mathcal{H}} . \tag{5}$$

I have already explained (in Section 2.4) why I consider the use of Dempster's rule in the first $n+1$ combinations of (5) acceptable; for the last one I have considered the general rule, because the assumption of the independence in an expression of such a generality is a nonsense. But this compels us to choose a JSSD with marginal beliefs $b^{pri} \oplus \bigoplus_{i=1}^{n} (h_i \rightharpoonup b_i)$ and b; by reasoning on the form of the belief $b^{pri} \oplus \bigoplus_{i=1}^{n} (h_i \rightharpoonup b_i)$ and considering that b is a belief on $\mathcal{U}$, it is not difficult to see that the choice of the JSSD is in a natural way completely determined by the choice of the JSSD's $m_{\mathcal{U}}^{\left(\biguplus_{i \in I} b_i\right) \otimes b}$ for $I \subset \{1, \ldots, n\}$ nonempty (the meaning of "natural" could be exactly explained, but this would take too much time, and I think that the intuitive meaning is clear). With such a reasoning, it is also easy to see that the only aspect of these JSSD's which has an influence on the posterior belief b^{post} is the conflict $c_I = c\left(\left(\biguplus_{i \in I} b_i\right) \otimes b\right)$; therefore b^{post} is completely determined by b^{pri} and the c_I's. Obviously, to be sensible, the chosen JSSD's cannot be completely independent: from the standpoint of the conflicts this can be expressed by saying that, for $I \cap J = \emptyset$, $c_{I \cup J}$ should be related to c_I and c_J. I think that the following constraints are sensible:

$$c_I + c_J - 1 \leq c_{I \cup J} \leq \min(c_I, c_J) . \tag{6}$$

The second inequality is clear: $\biguplus_{i \in I \cup J} b_i$ conflicts with b less than $\biguplus_{i \in I} b_i$ or $\biguplus_{i \in J} b_i$; whereas the first one can be understood by interpreting $1 - c_I$ as the "agreement" of $\biguplus_{i \in I} b_i$ with b: the agreement of $\biguplus_{i \in I \cup J} b_i$ with b is at most the sum of the agreements of $\biguplus_{i \in I} b_i$ and $\biguplus_{i \in J} b_i$ with b (since $1 - c_{I \cup J} \leq (1 - c_I) + (1 - c_J) \Leftrightarrow c_I + c_J - 1 \leq c_{I \cup J}$). For instance, using Dempster's rule, the conflicts $c_I = c\left(\left(\biguplus_{i \in I} b_i\right) \oplus b\right)$ satisfy the constraints (6), as is easy to show. However these constraints are not strictly necessary; I shall assume them only because this will simplify the expressions of the results. In particular, if $c_{\{i\}} = 0$ for every i, then $b^{post} = b^{pri}$; that is, only the conflict between the (scientific) hypotheses and the observations can make us change our beliefs: this is very much in the spirit of the philosophy of Popper (see for instance [24]). We shall see this better by analyzing the particular cases $n = 1$ and $n = 2$.

Consider the situation with a single hypothesis h_1: (5) becomes

$$b^{post} = \left(\left(b^{pri} \oplus (h_1 \rightharpoonup b_1)\right) \otimes b\right)\downarrow_{\{h_1\}} ;$$

and it is easy to see that

$$c = c\left(\left(b^{pri} \oplus (h_1 \rightharpoonup b_1)\right) \otimes b\right) = b^{pri}(h_1)\, c_{\{1\}},$$

in particular b^{post} is not defined only if we were sure of the correctness of the hypothesis and this one completely conflicts with the observations. If we are not in this situation, then

$$\begin{aligned} b^{post}(h_1) &= \frac{b^{pri}(h_1) - c}{1-c} = \frac{1 - c_{\{1\}}}{1-c} b^{pri}(h_1) \leq b^{pri}(h_1) \text{ and} \\ p^{b^{post}}(h_1) &= \frac{1 - c_{\{1\}}}{1-c} p^{b^{pri}}(h_1) \leq p^{b^{pri}}(h_1). \end{aligned}$$

Therefore we cannot increase our belief in a (scientific) hypothesis by confronting it with the observations: this is Hume's argument against the induction (or at least Popper's interpretation of this argument). The only way to increase the belief in a hypothesis is to consider some alternatives; in a moment I shall consider the case $n = 2$, but first I would like to formulate an important remark. In the situations in which $b^{pri}(h_1) \in (0,1)$, we maintain our belief in the correctness of h_1 if and only if $c_{\{1\}} = 0$; consequently, if we were using Dempster's rule to combine $b^{pri} \oplus (h_1 \rightharpoonup b_1)$ and b, then we could lose belief in the correctness of h_1 even in the case in which $b_1 = b$ (i.e., the prediction of the hypothesis is perfect), since $c(b \oplus b) = 0 \Leftrightarrow D(b) = 0$. It is probably to avoid these problems that Smets, who considers Dempster's one as the only conjunctive rule of combination, accepts only observations b which are certainties (the conflict of the conditioning is not criticizable); he has developed a formalism which would allow without problems any belief b, and he restricts them to certainties: this is suspicious. I think that in this situation more than ever the choice of a JSSD $m_{\mathcal{U}}^{b_1 \otimes b}$ which minimizes the conflict is sensible; since only the conflict $c_{\{1\}}$ influences the result b^{post}, if my Conjecture 55 were correct, then we would not need to find the JSSD minimizing the conflict to obtain the result.

Consider now the situation with two hypotheses h_1 and h_2: (5) becomes

$$b^{post} = \left(\left(b^{pri} \oplus (h_1 \rightharpoonup b_1) \oplus (h_2 \rightharpoonup b_2)\right) \otimes b\right)\downarrow_{\{h_1,h_2\}} ;$$

and it is easy to see that

$$\begin{aligned} c &= c\left(\left(b^{pri} \oplus (h_1 \rightharpoonup b_1) \oplus (h_2 \rightharpoonup b_2)\right) \otimes b\right) = \\ &= b^{pri}(h_1)\, c_{\{1\}} + b^{pri}(h_2)\, c_{\{2\}} + \left(b^{pri}((h_1 \vee h_2)) - b^{pri}(h_1) - b^{pri}(h_2)\right) c_{\{1,2\}}, \end{aligned}$$

in particular b^{post} can be not defined only if we were sure that the truth is among the hypotheses and at least one of the conflicts is total. If b^{post} is defined, then

$$\begin{aligned} b^{post}(h_1) &= \frac{b^{pri}(h_1) + (b^{pri}((h_1 \vee h_2)) - b^{pri}(h_1))\, c_{\{2\}} - c}{1-c} \quad \text{and} \\ p^{b^{post}}(h_1) &= \frac{1 - c_{\{1\}}}{1-c} p^{b^{pri}}(h_1); \end{aligned}$$

analogously for $b^{post}(h_2)$ and $p^{b^{post}}(h_2)$. Therefore both the belief and the plausibility that we assign to a hypothesis can increase, when the conflict of this hypothesis with the observations is smaller than the one of the alternative. As Popper says: a (scientific) hypothesis is accepted or rejected because it is better or worse than the other available (scientific) hypotheses, in the light of experimental observations. But

$$b^{post}((h_1 \vee h_2)) = \frac{b^{pri}((h_1 \vee h_2)) - c}{1-c}:$$

the belief in $(h_1 \vee h_2)$ behaves like the belief in h_1 in the case $n = 1$; and thus

$$b^{post}(h_1) \leq b^{post}((h_1 \vee h_2)) \leq b^{pri}((h_1 \vee h_2)):$$

the belief in a hypothesis will never be larger than the initial belief in the set of hypotheses. This is due to the fact that $(\neg h_1 \wedge \neg h_2)$ can be seen as an hypothesis which leads to the vacuous belief on $\mathcal{U}$, and $c\left(b_{\mathcal{U}}^{\emptyset} \otimes b\right) = 0$ (the JSSD is unique, since the vacuous belief is a certainty: the certainty of $\top$): thus, according to Popper, $(\neg h_1 \wedge \neg h_2)$ is not a scientific hypothesis, since it is not falsifiable. If we want a hypothesis to be believed without limitations, we must assume that the truth is among the hypotheses: $b^{pri}((h_1 \vee h_2)) = 1$; but in this case we can get debatable results, since $b^{post}((h_1 \vee h_2)) = 1$, independently of the conflicts with the observations. For instance, if $c_{\{2\}} = 1$ (the second hypothesis is incompatible with the observations), then $b^{post}(h_1) = 1$ independently of the prior belief and of $c_{\{1\}}$ (if b^{post} is defined): a very improbable hypothesis can become a certainty if the alternative is unacceptable. Another example: if (for a small $\varepsilon \neq 0$) $c_{\{1\}} = c_{\{2\}} = 1 - \varepsilon$, $c_{\{1,2\}} = 1 - 2\varepsilon$ and $b^{pri}(h_1) = b^{pri}(h_2) = 0$, then $b^{post}(h_1) = b^{post}(h_2) = \frac{1}{2}$, thus the belief in two very improbable hypothesis can grow a lot.

A last remark: considering that, when we compare some hypotheses with the results of some observations, the conflict between the respective beliefs is clearly a measure of disagreement, I have proposed to choose as the JSSD's $m_{\mathcal{U}}^{\left(\biguplus_{i \in I} b_i\right) \otimes b}$ the ones which minimize the conflict; thus it is important to notice that, if my Conjecture 55 were correct, then the conflicts c_I would satisfy the constraints (6):

$$\begin{aligned} &\max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_1(\varphi) - p^b(\varphi)\right) + \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_2(\varphi) - p^b(\varphi)\right) - 1 \leq \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left((b_1 \uplus b_2)(\varphi) - p^b(\varphi)\right) \text{ and} \\ &\max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left((b_1 \uplus b_2)(\varphi) - p^b(\varphi)\right) \leq \min\left(\max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_1(\varphi) - p^b(\varphi)\right), \max_{\varphi \in \mathcal{L}_{\mathcal{U}}} \left(b_2(\varphi) - p^b(\varphi)\right)\right). \end{aligned}$$

In fact $(b_1 \uplus b_2)(\varphi) - p^b(\varphi) = b_1(\varphi)\, b_2(\varphi) - p^b(\varphi) \leq b_1(\varphi) - p^b(\varphi)$ and analogously $(b_1 \uplus b_2)(\varphi) - p^b(\varphi) \leq b_2(\varphi) - p^b(\varphi)$; whereas

$$\begin{aligned} (b_1 \uplus b_2)((\varphi \vee \psi)) - p^b((\varphi \vee \psi)) &\geq b_1(\varphi)\, b_2(\psi) - p^b(\varphi) - p^b(\psi) \geq \\ &\geq b_1(\varphi) + b_2(\psi) - 1 - p^b(\varphi) - p^b(\psi)\,. \end{aligned}$$

References

[1] Salem Benfehrat, Alessandro Saffiotti and Philippe Smets, **Belief Functions and Default Reasoning**, Technical Report TR/IRIDIA/95-05, Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle (IRIDIA), Université Libre de Bruxelles, 1995

[2] Gerhard Brewka, Jürgen Dix and Kurt Konolige, *A Tutorial on Nonmonotonic Reasoning*, in: G. Brewka, K. P. Jantke and P. H. Schmitt eds., **Nonmonotonic and Inductive Logic, Lecture Notes in Artificial Intelligence 659**, pp. 1-88, Springer-Verlag, 1993

[3] Bruno de Finetti, *Foresight: Its Logical Laws, Its Subjective Sources*, in: H. E. Kyburg and H. E. Smokler eds., **Studies in Subjective Probability**, pp. 53-118, 2nd edition, Krieger Publishing, 1980

(translation of: La prévision: ses lois logiques, ses sources subjectives, Annales de l'Institut Henri Poincaré, 7, pp. 1-68, 1937)

[4] Arthur P. Dempster, *Upper and Lower Probabilities Induced by a Multivalued Mapping*, **Annals of Mathematical Statistics, 38**, pp. 325-339, 1967

[5] Arthur P. Dempster, *A Generalization of Bayesian Inference*, **Journal of the Royal Statistical Society, B 30**, pp. 205-232, 1968

[6] Persi Diaconis, *Review of "A Mathematical Theory of Evidence"*, **Journal of the American Statistical Society, 73**, pp. 677-678, 1978

[7] Didier Dubois and Henri Prade, **Possibility Theory, An Approach to Computerized Processing of Uncertainty**, Plenum Press, 1988

(slightly revised translation of: Théorie des possibilités, Applications à la représentation des connaissances en informatique, Masson, 1985)

[8] Didier Dubois and Henri Prade, *On the Unicity of Dempster Rule of Combination*, **International Journal of Intelligent Systems, 1**, pp. 133-142, 1986

[9] Didier Dubois and Henri Prade, *Consonant Approximations of Belief Functions*, **International Journal of Approximate Reasoning, 4**, pp. 419-449, 1990

[10] Didier Dubois and Henri Prade, *Focusing versus Updating in Belief Function Theory*, in: R. R. Yager, J. Kacprzyk and M. Fedrizzi, **Advances in the Dempster-Shafer Theory of Evidence**, pp. 71-95, John Wiley & Sons, 1994

[11] Richard L. Epstein, **The Semantic Foundations of Logic, Volume 1: Propositional Logics**, Kluwer Academic Publishers, 1990

[12] Ronald Fagin and Joseph Y. Halpern, *A New Approach to Updating Beliefs*, in: P. P. Bonissone, M. Hendrion, L. N. Kanal and J. F. Lemmer eds., **Uncertainty in Artificial Intelligence 6, Machine Intelligence and Pattern Recognition, 12**, pp. 347-374, North-Holland, 1991

[13] Carl G. Hempel, *Deductive-Nomological vs. Statistical Explanation*, in: H. Feigl and G. Maxwell eds., **Minnesota Studies in the Philosophy of Science, Volume III: Scientific Explanation, Space, and Time**, pp. 98-169, University of Minnesota Press, 1962

[14] Carl G. Hempel, **Aspects of Scientific Explanation**, The Free Press, 1965

[15] Frank Klawonn and Erhard Schwecke, *On the Axiomatic Justification of Dempster's Rule of Combination*, **International Journal of Intelligent Systems, 7**, pp. 469-478, 1992

[16] Stephen C. Kleene, **Mathematical Logic**, John Wiley & Sons, 1967

[17] George J. Klir, *Measures of Uncertainty in the Dempster-Shafer Theory of Evidence*, in: R. R. Yager, J. Kacprzyk and M. Fedrizzi, **Advances in the Dempster-Shafer Theory of Evidence**, pp. 35-49, John Wiley & Sons, 1994

[18] Henry E. Kyburg Jr., **Probability and the Logic of Rational Belief**, Wesleyan University Press, 1961

[19] Henry E. Kyburg Jr., *Bayesian and Non-Bayesian Evidential Updating*, **Artificial Intelligence, 31**, pp. 271-293, 1987

[20] Wiktor Marek and Mirosław Truszczyński, **Nonmonotonic Logic, Context-Dependent Reasoning**, Springer-Verlag, 1993

[21] Halina Mortimer, **The Logic of Induction**, Hellis Horwood Limited, 1988

(translation, with additional material, of: Logika indukji, Panstwowe Wydawnictwo Naukowe, 1982)

[22] Frederick Mosteller, **Fifty Challenging Problems in Probability with Solutions**, Addison-Wesley, 1965

[23] Nikhil R. Pal, James C. Bezdek and Rohan Hemasinha, *Uncertainty Measures for Evidential Reasoning I: A Review*, **International Journal of Approximate Reasoning, 7**, pp. 165-183, 1992

[24] Karl R. Popper, *Realism and the Aim of Science*, Volume 1 of: W. W. Bartley III ed., **Postscript to the Logic of Scientific Discovery**, Hutchinson, 1983

[25] Arthur Ramer, *Fuzzy Information and Combinatorial Inequalities*, **Kybernetika, 28**, pp. 4-11 of the supplement (Proceedings of the International Symposium on Fuzzy Approach to Reasoning and Decision Making, Bechyně, Czechoslovakia, 1990), 1992

[26] Glenn Shafer, **A Mathematical Theory of Evidence**, Princeton University Press, 1976

[27] Glenn Shafer, *Non-additive Probabilities in the Work of Bernoulli and Lambert*, **Archive for History of Exact Sciences, 19**, pp. 309-370, 1978

[28] Glenn Shafer, *The Combination of Evidence*, **International Journal of Intelligent Systems, 1**, pp. 155-179, 1986

[29] Glenn Shafer, *Perspectives on the Theory and Practice of Belief Functions*, **International Journal of Approximate Reasoning, 4**, pp. 323-362, 1990

[30] Philippe Smets, *Belief Functions*, in: P. Smets, A. Mamdani, D. Dubois and H. Prade eds., **Non-Standard Logics for Automated Reasoning**, pp. 253-277, Academic Press, 1988

[31] Philippe Smets, *The Combination of Evidence in the Transferable Belief Model*, **The Institute of Electrical and Electronics Engineers (IEEE) Transactions on Pattern Analysis and Machine Intelligence (PAMI), 12**, pp. 447-458, 1990

[32] Philippe Smets, *Belief Functions: the Disjunctive Rule of Combination and the Generalized Bayesian Theorem*, **International Journal of Approximate Reasoning, 9**, pp. 1-35, 1993

[33] Philippe Smets, *The α-junctions: Combination Operators Applicable to Belief Functions*, in: D. M. Gabbay, R. Kruse, A. Nonnengart and H. J. Ohlbach eds., **Qualitative and Quantitative Practical Reasoning**, pp. 131-153, Springer-Verlag, 1997

[34] Ronald R. Yager, *On Considerations of Credibility of Evidence*, **International Journal of Approximate Reasoning, 7**, pp. 45-72, 1992

[35] Ronald R. Yager, *Nonmonotonic Reasoning with Belief Structures*, in: R. R. Yager, J. Kacprzyk and M. Fedrizzi, **Advances in the Dempster-Shafer Theory of Evidence**, pp. 533-554, John Wiley & Sons, 1994