# Chapter 4

# Dempster–Shafer Belief

In this chapter we shall give two arguments in support of belief functions being identified (under certain conditions) with Dempster–Shafer (DS) belief functions, originally introduced by Dempster in [11], and will relate them to probability functions. First however we shall develop a little of their theory. A much fuller account of these functions may be found in Shafer's seminal treatise [63]. Our framework may appear slightly more restrictive than that used by Shafer, but the differences are, in fact, inconsequential.

## An Equivalent Definition of DS-belief Functions

The definition of DS-belief Functions given in Chapter 1 was chosen there because it required no additional notation and because it was easy to see a relationship with belief as probability. However an alternative, equivalent, definition which we now give is, in practice, easier to work with and more transparent.

*Alternative definition: Bel:* $SL \longrightarrow [0,1]$ *is a Dempster–Shafer belief function if there is a function* $m : \overline{SL} \longrightarrow [0,1]$ *such that*

$$\sum_{\overline{\theta} \in \overline{SL}} m(\overline{\theta}) = 1, \quad m(\mathbf{0}) = 0$$

*and for all* $\phi \in SL$,

$$Bel\,(\phi) = \sum_{\overline{\theta} \leq \overline{\phi}} m(\overline{\theta}).$$

Here $\overline{SL}$ is the Lindenbaum algebra of $L$, that is $\overline{SL} = \{\overline{\theta} \mid \theta \in SL\}$ where $\overline{\theta} = \{\phi \in SL \mid \theta \equiv \phi\}$ and $\equiv$ is the equivalence relation $\models (\theta \leftrightarrow \phi)$, with the (well defined) operations, constants and relations

$$\overline{\theta} \wedge \overline{\phi} = \overline{(\theta \wedge \phi)}, \quad \overline{\theta} \vee \overline{\phi} = \overline{(\theta \vee \phi)}, \quad \neg\overline{\theta} = \overline{(\neg\theta)},$$

$$\mathbf{1} = \overline{\theta \vee \neg\theta}, \quad \mathbf{0} = \overline{\theta \wedge \neg\theta},$$

$$\overline{\theta} \leq \overline{\phi} \iff \models (\theta \rightarrow \phi) \iff \overline{\theta} \wedge \overline{\phi} = \overline{\theta}.$$

It is useful to observe that $\overline{\theta} \mapsto S_\theta$ is an isomorphism of this algebra with the field of all subsets of $At^L$.

The function $m$ in this definition is called a basic probability assignment, bpa, for $Bel$. Notice that according to this definition belief in $\phi$ is a sum of 'basic chunks of belief', $m(\overline{\theta})$, in the $\overline{\theta} \leq \overline{\phi}$, equivalently in the $S_\theta$ for $S_\theta \subseteq S_\phi$. In this sense then $m(\overline{\theta})$ is the belief in $\theta$ beyond that in any $\psi$ with $\overline{\psi} < \overline{\theta}$.

The equivalence of these definitions follows from the next two results due to Shafer.

**Theorem 4.1** *Let $Bel^m$ be the DS-belief function defined via the bpa $m$ as in the above definition. Then $Bel^m$ satisfies* $(DS1\text{–}3)$.

**Proof** $(DS1)$ and $(DS2)$ are immediate since if $\models (\theta \leftrightarrow \phi)$ then $\overline{\theta} = \overline{\phi}$ and if $\models \theta$ then $\overline{\theta} = \mathbf{1}, \overline{\neg\theta} = \mathbf{0}$ so $\overline{\phi} \leq \overline{\theta}$ for all $\overline{\phi}$ and

$$Bel^m(\theta) = \sum_{\overline{\phi}} m(\overline{\phi}) = 1, \quad Bel^m(\neg\theta) = \sum_{\overline{\phi} \leq \mathbf{0}} m(\overline{\phi}) = m(\mathbf{0}) = 0.$$

It only remains to prove $(DS3)$. Using the notation of $(DS3)$,

$$\sum_{\emptyset \neq S} (-1)^{|S|-1} Bel^m(\bigwedge_{i \in S} \theta_i) = \sum_{\emptyset \neq S} (-1)^{|S|-1} \sum_{\overline{\psi} \leq \overline{\bigwedge_{i \in S} \theta_i}} m(\overline{\psi})$$

$$= \sum_{I(\overline{\psi}) \neq \emptyset} m(\overline{\psi}) \sum_{\emptyset \neq S \subseteq I(\overline{\psi})} (-1)^{|S|-1} \quad \text{where } I(\overline{\psi}) = \{i \mid \overline{\psi} \leq \overline{\theta}_i\}$$

$$= \sum_{I(\overline{\psi}) \neq \emptyset} m(\overline{\psi}) = \sum_{\substack{\overline{\psi} \leq \overline{\theta}_i \\ \text{for some } i}} m(\overline{\psi}) \leq \sum_{\overline{\psi} \leq \overline{\bigvee_i \theta_i}} m(\overline{\psi}) = Bel^m(\bigvee_i \theta_i),$$

as required, since for a finite set $X \neq \emptyset$,

$$\begin{aligned} \sum_{Y \subseteq X} (-1)^{|Y|-1} &= \text{Number of odd cardinality subsets of } X \\ &\quad - \text{Number of even cardinality subsets of } X \\ &= 0 \quad \text{by induction on } |X|, \end{aligned}$$

so

$$\sum_{\substack{\emptyset \neq S \\ S \subseteq I(\overline{\psi})}} (-1)^{|S|-1} = 0 - (-1)^{|\emptyset|-1} = 1 \text{ for } I(\overline{\psi}) \neq \emptyset.$$

□

**Theorem 4.2** *If Bel satisfies (DS1–3) then there is a unique bpa m such that Bel = $Bel^m$, where $Bel^m$ is defined from m as in the above definition.*

**Proof** Set $m(\mathbf{0}) = 0$ and for $\emptyset \neq R = \{\alpha_{j_1}, ..., \alpha_{j_q}\} \subseteq \{\alpha_1, ..., \alpha_J\}$ set

$$m(\overline{\bigvee R}) = \sum_{S \subseteq R} (-1)^{|R-S|} \, Bel(\bigvee S).$$

We first show that $m(\overline{\bigvee R}) \geq 0$. This is clear if $| R | \leq 1$. Otherwise

$$m(\overline{\bigvee R}) = Bel(\bigvee R) + \sum_{S \subset R} (-1)^{|R-S|} \, Bel(\theta_{i_1} \wedge ... \wedge \theta_{i_k}),$$

by $(DS2)$, where $R - S = \{\alpha_{j_{i_1}}, ..., \alpha_{j_{i_k}}\}$ and $\theta_{i_r} = \alpha_{j_1} \vee \alpha_{j_2} \vee ... \vee \alpha_{j_{i_r - 1}} \vee \alpha_{j_{i_r + 1}} \vee ... \vee \alpha_{j_q}$. This further equals

$$Bel(\bigvee_{i=1}^{q} \theta_i) + \sum_{\emptyset \neq T \subseteq \{1,...,q\}} (-1)^T \, Bel\,(\bigwedge_{i \in T} \theta_i) \geq 0 \;\; by \;(DS3).$$

Hence, since $\overline{\phi} = \overline{\bigvee S_\phi}$, $m(\overline{\phi}) \geq 0$ for all $\overline{\phi} \in \overline{SL}$. Also

$$\begin{aligned}
\sum_{\overline{\theta} \leq \overline{\phi}} m(\overline{\theta}) &= \sum_{R \subseteq S_\phi} m(\overline{\bigvee R}) \\
&= \sum_{R \subseteq S_\phi} \sum_{S \subseteq R} (-1)^{|R-S|} \, Bel(\bigvee S) \\
&= \sum_{S} Bel(\bigvee S) \sum_{S \subseteq R \subseteq S_\phi} (-1)^{|R-S|} = Bel(\bigvee S_\phi) = Bel(\phi)
\end{aligned}$$

by $(DS2)$ since, as above, the sum $\sum_{S \subseteq R \subseteq S_\phi} (-1)^{|R-S|}$ is zero if $S \subset S_\phi$. Hence also

$$\sum_{\overline{\theta}} m(\overline{\theta}) = \sum_{\overline{\theta} \leq \overline{\phi \vee \neg \phi}} m(\overline{\theta}) = Bel(\phi \vee \neg\phi) = 1,$$

so $m$ is a bpa and $Bel = Bel^m$. Finally $m$ is the unique bpa for $Bel$ since if $m, m'$ are distinct bpa's then there must be a smallest set $R \subseteq \{\alpha_1, ..., \alpha_J\}$ for which $m(\overline{\bigvee R}) \neq m'(\overline{\bigvee R})$ and so $Bel^m(\bigvee R) \neq Bel^{m'}(\bigvee R)$ □

We now give a justification (or explanation) of Dempster–Shafer belief. A second justification will be given later when we have developed a little more notation and familiarity.

## First Justification

Consider the following situation. Suppose an agent knows he will be receiving a message of the form '$\theta$ is true' for some $\theta \in SL$, $\theta$ consistent. For example the message might arise from some experiment. Let $p(\overline{\theta})$ be the, possibly subjective, probability that the message will be that '$\theta$ is true' (or a sentence logically equivalent to $\theta$, which the agent would take to be just as good). Then $\sum_{\overline{\theta}} p(\overline{\theta}) = 1$, $p(\overline{\theta}) \geq 0$, $p(\mathbf{0}) = 0$ and for $\phi \in SL$,

$$\begin{array}{lcl} \text{Probability agent} & = & \text{Probability he will receive a message} \\ \text{will learn } \phi \text{ is true} & & \text{`}\theta\text{ is true' with } \theta \text{ logically implying } \phi \\ & & \\ & = & \sum_{\overline{\theta} \leq \overline{\phi}} p(\overline{\theta}). \end{array}$$

Thus, according to the definition above, identifying belief in $\phi$ with the probability of *learning* the truth of $\phi$ gives belief as a DS-belief function and at the same time provides an explanation of the bpa. From the point of view of belief as probability a DS-belief function is measuring the belief that one will *learn* the truth of a sentence $\theta$ rather than just the truth of $\theta$.

Whether or not that is what an expert means when he gives figures such as those in our example $\mathbb{E}$ is perhaps debatable, although the above example shows how sets of constraints $K$ might arise in which Dempster–Shafer belief was appropriate.

A second criticism one might raise against this justification is that, whilst learning the truth of a sentence logically equivalent to $\theta$ might *theoretically* be just as good as learning $\theta$, the *practical* problem of deciding (in general) whether such a logical equivalence holds is, assuming $P \neq NP$, infeasible. (A similar criticism can be raised at various other points in this book.)

A particular example of this situation occurs by considering a refinement of the urn model used to justify belief as probability in the previous chapter. Recall that, in the case of a doctor diagnosing patients, the idea was that the doctor's belief in $\theta$, $Bel(\theta)$, is identified with

$$\frac{|\{x \in M \mid x \text{ has } \theta\}|}{|M|}$$

where $M$ is the set of patients $x$ he has previously seen. Here it was assumed that the doctor has complete knowledge of each previous patient $x$, that is that $x$ having $\theta$ is the same as the doctor knowing that $x$ has $\theta$.

However, this is clearly unrealistic: in practice the knowledge, $\phi_x$, the doctor has about $x$ would not necessarily be complete, i.e. $\phi_x$ would not necessarily be an atom. Nevertheless if we now proceed as before to define $Bel_0$ by

$$\begin{aligned} Bel_0(\theta) &= \frac{|\{x \in M \mid \text{ The doctor knows } x \text{ has } \theta\}|}{|M|} \\ &= \frac{|\{x \in M \mid \ \phi_x \models \theta\}|}{|M|}, \end{aligned}$$

then it is easy to check that $Bel_0$ is a DS-belief function with bpa

$$m(\overline{\theta}) = \frac{|\{x \in M \mid \overline{\phi_x} = \overline{\theta}\}|}{|M|}.$$

Again it seems questionable whether the figures given by the doctor in our example $\mathbb{E}$ could be interpreted in this way. However this clearly provides an argument that DS-belief is a possibility for an intelligent agent.

## Plausibility

Directly from $(DS3)$ we see that for $Bel$ a DS-belief function and $\theta, \phi \in SL$,

$$Bel(\theta \vee \phi) \geq \ Bel(\theta) + \mathrm{Bel}(\phi) - \ Bel(\theta \wedge \phi).$$

In particular for $\phi = \neg\theta$,

$$1 = \ Bel(\theta \vee \neg\theta) \geq \ Bel(\theta) + \ Bel(\neg\theta) \quad \text{since} \quad Bel(\theta \wedge \neg\theta) = 0$$

so $Bel(\theta) \leq 1 - \ Bel(\neg\theta)$.

The difference $1 - (Bel(\theta) + \ Bel(\neg\theta))$ could be thought of as the unassigned, or uncommitted, belief between $\theta$ and $\neg\theta$, a large value here corresponding to ignorance of $\theta$ and $\neg\theta$. In particular complete ignorance of $\theta$ and $\neg\theta$ would correspond to $Bel(\theta) = \ Bel(\neg\theta) = 0$, as opposed to $Bel(\theta) = \ Bel(\neg\theta) = \frac{1}{2}$ for belief as probability. In this way then DS-belief functions could be said to have the ability to distinguish between genuine uncertainty and simple ignorance.

The plausibility of $\theta$, $Pl(\theta)$, is defined by

$$Pl(\theta) = 1 - \ Bel(\neg\theta)$$

and could be thought of as unassigned belief which could all go to $\theta$. Notice that $Bel(\theta) \leq 1 - \ Bel(\neg\theta) = Pl(\theta)$ and in terms of the bpa $m$ for $Bel$,

$$Pl(\theta) = 1 - Bel(\neg\theta) = \sum_{\overline{\phi}} m(\overline{\phi}) - \sum_{\overline{\phi} \leq \overline{\neg\theta}} m(\overline{\phi}) = \sum_{\mathbf{0} < \overline{\phi} \wedge \overline{\theta}} m(\overline{\phi}).$$

(Many other perceived facets of belief can similarly be captured by DS-belief functions but they will not be relevant here.)

## A Representation of DS-belief Functions

Let $\overline{\theta}_i, i = 1, ..., r$ enumerate the non-zero elements of $\overline{SL}$ (so $r = 2^{2^n} - 1$ for our 'default' language $L = \{p_1, ..., p_n\}$). Then the above theorem 4.2 shows that any DS-belief function $Bel$ on $SL$ can be uniquely specified by the vector

$$\langle m(\overline{\theta}_1), ..., m(\overline{\theta}_r)\rangle$$

where $m$ is the (unique) bpa of $Bel$. Furthermore, of course, the $m(\overline{\theta}_i) \geq 0$ and $\sum_i m(\overline{\theta}_i) = 1$.

Conversely given a vector $\langle x_1, ..., x_r\rangle$ with $x_i \geq 0$, $\sum_i x_i = 1$ we can define a bpa $m$ by $m(\overline{\theta}_i) = x_i$, $m(\mathbf{0}) = 0$ and hence a DS-belief function corresponding to this vector.

To sum up then there is a 1–1 correspondence between DS-belief functions on SL and points in

$$\{\langle x_1, ..., x_r\rangle \mid \sum x_i = 1, x_i \geq 0\}.$$

This is exactly similar to the situation for probability functions except that now the vector is of length $2^{2^n} - 1$ rather than $2^n$.

Again, exactly as for belief as probability, we can identify DS-belief functions satisfying $K$ (as given at the close of Chapter 1) with points $\langle x_1, ..., x_r\rangle \in \mathbb{R}^r$ satisfying

$$\vec{x}D_k = \vec{e_k}, \quad \vec{x} \geq 0$$

for some matrix $D_k$ and vector $\vec{e_k}$ (obtained by replacing each $Bel(\theta_j)$ by $\sum_{\overline{\phi} \leq \overline{\theta}_j} m(\overline{\phi})$ throughout $K$ and adding in also $\sum_{\overline{\phi}} m(\overline{\phi}) = 1$). However, the now double exponential length of $\vec{x}$ makes using this form of the question $Q$ somewhat more complicated in practice.

## Probability and DS-belief

Using the first definition of DS-belief functions it is clear that any probability function is also a DS-belief function since, by proposition 2.1 and ($P1$), ($DS1$–3) hold, with equality, for any probability function. The converse is false however. To see this notice that if the bpa $m$ gives a probability function $Bel$ then

$$1 = \sum_{i=1}^{J} Bel(\alpha_i) = \sum_{i=1}^{J} \sum_{\overline{\phi} \leq \overline{\alpha}_i} m(\overline{\phi}) = \sum_{i=1}^{J} (m(\mathbf{0}) + m(\overline{\alpha}_i)) = \sum_{i=1}^{J} m(\overline{\alpha}_i),$$

and conversely if $\sum_{i=1}^{J} m(\overline{\alpha}_i) = 1$ then for $\overline{\phi} \notin \{\overline{\alpha}_i \mid i = 1, ..., J\}$, $m(\overline{\phi}) = 0$, since $\sum_{\overline{\phi}} m(\overline{\phi}) = 1$ and $m(\overline{\phi}) \geq 0$, so

$$Bel(\theta) = \sum_{\overline{\phi} \leq \overline{\theta}} m(\overline{\phi}) = \sum_{\overline{\alpha}_i \leq \overline{\theta}} m(\overline{\alpha}_i) = \sum_{\alpha_i \in S_\theta} Bel(\alpha_i),$$

from which it is clear from Chapter 1 that $Bel$ is a probability function. Thus a DS-belief function is a probability function just if $\sum_{i=1}^{J} m(\overline{\alpha}_i) = 1$ and hence to produce a DS-belief function which is not a probability function it is enough to choose the bpa $m$ such that $m(\overline{\phi}) > 0$ for some $\overline{\phi} \notin \{\overline{\alpha}_i \mid i = 1, ..., J\}$.

A second equivalent to a DS-belief function $Bel$ being a probability function is that

$$Bel(\theta) + Bel(\neg\theta) = 1, \quad i.e.\ \ Bel(\theta) = Pl(\theta), \quad \text{for all } \theta \in SL.$$

For clearly this condition holds if $Bel$ is a probability function. Conversely if $Bel$ is not a probability function suppose $m(\overline{\phi}) > 0$ where $\overline{\phi} \notin \{\overline{\alpha}_i \mid i = 1, ..., J\}$ and $m$ is the bpa of $Bel$. Then for $\alpha_j \in S_\phi$ ($\neq \emptyset$ since $m(\overline{\phi}) > 0$) there are no non-zero common terms in the sums

$$\sum_{\overline{\theta} \leq \overline{\alpha}_j} m(\overline{\theta}), \quad \sum_{\overline{\theta} \leq \overline{\neg\alpha_j}} m(\overline{\theta})$$

and neither sum contains $m(\overline{\phi})$, since $\overline{\alpha}_j < \overline{\phi}$, so

$$Bel(\alpha_j) + Bel(\neg\alpha_j) < \sum_{\overline{\theta}} m(\overline{\theta}) = 1.$$

Another connection between probability and DS-belief is provided by the following theorem due to Dempster [11] and Kyburg [41] which shows that DS-belief functions can be viewed as *sets* of probability functions. Alternatively this theorem can be viewed as showing to what extent a DS-belief function can be 'refined' into a probability function.

In the following theorem let $Bel$ be a DS-belief function on $SL$ and let

$$W(Bel) = \{w \mid \ w \text{ is a probability function on } SL \text{ and } w(\theta) \geq Bel(\theta) \text{ for all } \theta \in SL\}.$$

**Theorem 4.3** *(i) If Bel is a probability function then* $W(Bel) = \{Bel\}$.

*(ii) For each* $\theta \in SL$ *there is* $w \in W(Bel)$ *such that* $w(\theta) = Bel(\theta)$.

*(iii) For each* $\theta \in SL$ *and* $w \in W(Bel)$, $w(\theta) \leq 1 - Bel(\neg\theta) = Pl(\theta)$.

*(iv) For each $\theta \in SL$ there is $w \in W(Bel)$ such that $w(\theta) = 1 - Bel(\neg\theta) = Pl(\theta)$.*

**Proof** For (iii) notice that if $w(\theta) > 1 - Bel(\neg\theta)$ then

$$Bel(\neg\theta) > 1 - w(\theta) = w(\neg\theta)$$

contradicting $w \in W(Bel)$.

For (i) notice that if $Bel$ is a probability function then $Bel \in W(Bel)$ and if $w \in W(Bel)$ and $w \neq Bel$ then $w(\theta) > Bel(\theta)$ for some $\theta$ so

$$w(\neg\theta) = 1 - w(\theta) < 1 - Bel(\theta) = Bel(\neg\theta)$$

contradicting $w \in W(Bel)$.

Part (iv) follows from (ii) since if $w(\neg\theta) = Bel(\neg\theta)$ with $w \in W(Bel)$ then $w(\theta) = 1 - w(\neg\theta) = 1 - Bel(\neg\theta)$.

So it only remains to prove (ii). Given $\theta \in SL$ let $m$ be the bpa of $Bel$. For each $\overline{\psi} > \mathbf{0}$ pick an atom $\beta_{\overline{\psi}} \in S_\psi$ such that if $S_{\psi\wedge\neg\theta} \neq \emptyset$ then $\beta_{\overline{\psi}} \in S_{\psi\wedge\neg\theta} \subseteq S_\psi$. Now define, for each atom $\alpha$, $w(\alpha) = \sum_{\beta_{\overline{\psi}}=\alpha} m(\overline{\psi}) \geq 0$. Clearly each $m(\overline{\psi})$ is associated with exactly one atom $\alpha$ so

$$\sum_\alpha w(\alpha) = \sum_{\overline{\psi}} m(\overline{\psi}) = 1$$

and hence $w$ extends to a probability function by defining

$$w(\theta) = \sum_{\alpha\in S_\theta} w(\alpha) \ for\ \theta \in SL.$$

For any $\overline{\phi}$,

$$\begin{aligned} Bel(\phi) &= \sum_{\overline{\psi}\leq\overline{\phi}} m(\overline{\psi}) = \sum_{\alpha\in S_\phi} \sum_{\substack{\overline{\psi}\leq\overline{\phi} \\ \beta_{\overline{\psi}}=\alpha}} m(\overline{\psi}) \leq \sum_{\alpha\in S_\phi} \sum_{\beta_{\overline{\psi}}=\alpha} m(\overline{\psi}) = \sum_{\alpha\in S_\phi} w(\alpha) \\ &= w(\phi) \end{aligned}$$

so $w \in W(Bel)$. Finally in the case $\phi = \theta$, if $\beta_{\overline{\psi}} = \alpha \in S_\phi$ it must be the case that, by choice of $\beta_{\overline{\psi}}$, $S_{\psi\wedge\neg\theta} = \emptyset$, i.e. $\overline{\psi} \leq \overline{\theta}$, so that in the above expression equality holds and $Bel(\theta) = w(\theta)$. □

This result is attractive in the sense that it identifies a DS-belief function $Bel$ with the (consistent) *set* of inequality constraints

$$w(\theta) \geq Bel(\theta) \quad \text{for} \quad \theta \in SL,$$

on a probability function $w$. Unfortunately not all such (consistent) sets of constraints correspond to DS-belief functions.

To see this consider, for $L = \{p, q\}$, the set of inequality constraints

$$w(\alpha_i) \geq 0,\ w(\alpha_i \vee \alpha_j) \geq \frac{1}{3},\ w(\alpha_i \vee \alpha_j \vee \alpha_k) \geq \frac{2}{3},\ w(\alpha_1 \vee \alpha_2 \vee \alpha_3 \vee \alpha_4) \geq 1$$

for distinct $i, j, k \in \{1, ..., 4\}$. For each of these lower bounds there is a probability function satisfying the constraints and assuming that lower bound. So if *Bel* was a DS-belief function such that $W(Bel)$ was exactly the set of (probability function) solutions to these constraints we should have, by theorem 4.3 (ii), that

$$Bel(\alpha_i) = 0, \quad Bel(\alpha_i \vee \alpha_j) = \frac{1}{3}, \quad Bel(\alpha_i \vee \alpha_j \vee \alpha_k) = \frac{2}{3}$$

for distinct $i, j, k \in \{1, ..., 4\}$. But for the bpa, $m$, of *Bel* this forces $m(\overline{\alpha_i}) = 0$, $m(\overline{\alpha_i \vee \alpha_j}) = \frac{1}{3}$ so

$$\frac{2}{3} = Bel(\alpha_1 \vee \alpha_2 \vee \alpha_3) \geq m(\overline{\alpha_1 \vee \alpha_2}) + m(\overline{\alpha_1 \vee \alpha_3}) + m(\overline{\alpha_2 \vee \alpha_3}) = 1,$$

contradiction.

## Conditional DS-belief

Amongst the several possible alternatives which have been suggested the (currently) most popular way to define a conditional DS-belief function from a DS-belief function *Bel* is to set

$$Bel(\theta \mid \phi) = \frac{Bel(\theta \vee \neg\phi) - Bel(\neg\phi)}{1 - Bel(\neg\phi)} \quad \text{whenever } Bel(\neg\phi) \neq 1.$$

This rather unlikely looking formula arises as follows. According to Shafer if $m_1, m_2$ are bpa's for DS-belief functions corresponding to 'independent' sources of belief then they may be combined using *Dempster's rule of combination* to give a bpa

$$m_1 \oplus m_2(\overline{\psi}) = \frac{\sum_{\overline{\lambda} \wedge \overline{\tau} = \overline{\psi}} m_1(\overline{\lambda}) m_2(\overline{\tau})}{1 - \sum_{\overline{\lambda} \wedge \overline{\tau} = \mathbf{0}} m_1(\overline{\lambda}) m_2(\overline{\tau})} \quad \text{for} \quad \overline{\psi} > \mathbf{0}$$

which is well defined provided the denominator is non-zero. That this is indeed a bpa in this case follows by noticing that

$$1 = \sum_{\overline{\lambda}} m_1(\overline{\lambda}) \cdot \sum_{\overline{\tau}} m_2(\overline{\tau}) = \sum_{\overline{\psi}} \sum_{\overline{\lambda} \wedge \overline{\tau} = \overline{\psi}} m_1(\overline{\lambda}) m_2(\overline{\tau}).$$

If the denominator is zero it must be the case that whenever $m_1(\overline{\lambda}), m_2(\overline{\tau}) > 0$ then $\overline{\lambda} \wedge \overline{\tau} = \mathbf{0}$. In this case then $Bel_1(\theta) = 1 = Bel_2(\phi)$, where $\overline{\theta} = \bigvee_{m_1(\overline{\lambda})>0} \overline{\lambda}$, $\overline{\phi} = \bigvee_{m_2(\overline{\tau})>0} \overline{\tau}$ , and $\theta, \phi$ contradict each other. Clearly in this circumstance it seems unreasonable to expect to be able to combine $Bel_1$ and $Bel_2$.

Now suppose that $m$ is the bpa of $Bel$ and we wish to condition on $\phi$. In this context it seems reasonable to represent the 'evidence' that $\phi$ holds by the DS-belief function with bpa $m'$ defined by $m'(\overline{\phi}) = 1$, $m'(\overline{\lambda}) = 0$ for $\overline{\lambda} \neq \overline{\phi}$ (obviously we cannot do this if $\phi$ is contradictory, i.e. $\overline{\phi} = \mathbf{0}$) and to take $Bel$ conditioned on $\phi$ to be the DS-belief function with bpa $m \oplus m'$. This gives

$$Bel(\theta \mid \phi) = \sum_{\overline{\psi} \leq \overline{\theta}} m \oplus m'(\overline{\psi}) = k \sum_{\mathbf{0} < \overline{\psi} \leq \overline{\theta}} \; \sum_{\overline{\lambda \wedge \phi} = \overline{\psi}} m(\overline{\lambda})$$

where $k = (1 - \sum_{\overline{\lambda} \wedge \overline{\phi} = \mathbf{0}} m(\overline{\lambda}))^{-1} = (1 - Bel(\neg\phi))^{-1}$.

Hence

$$Bel(\theta \mid \phi) = k \sum_{\mathbf{0} < \overline{\lambda} \wedge \overline{\phi} \leq \overline{\theta}} m(\overline{\lambda}) = k \left( \sum_{\overline{\lambda} \leq \overline{\theta} \vee \overline{\neg\phi}} m(\overline{\lambda}) - \sum_{\overline{\lambda} \leq \overline{\neg\phi}} m(\overline{\lambda}) \right)$$

$$= \frac{Bel(\theta \vee \neg\phi) - Bel(\neg\phi)}{1 - Bel(\neg\phi)},$$

provided the denominator is non-zero, i.e. provided $Bel$ does not give belief 1 to $\neg\phi$.

Notice that for the corresponding plausibilities we obtain the expression

$$Pl(\theta \mid \phi) = 1 - Bel(\neg\theta \mid \phi) = 1 - \frac{Bel(\neg\theta \vee \neg\phi) - Bel(\neg\phi)}{1 - Bel(\neg\phi)}$$

$$= \frac{1 - Bel(\neg\theta \vee \neg\phi)}{1 - Bel(\neg\phi)} = \frac{Pl(\theta \wedge \phi)}{Pl(\phi)}$$

which is exactly similar to that for conditional probability.

In connection with the justification for reducing the set of constraints $K$ in question $Q$ (at the close of Chapter 1) to involve simply unconditional belief, notice that if $1 - Bel(\neg\phi) = 0$ then also $Bel(\theta \vee \neg\phi) - Bel(\neg\phi) = 0$ for a DS-belief function $Bel$, so that, as for probability, the unconditional form of the constraint formed by multiplying through by the denominator of the $Bel(\theta_i \mid \phi)$ will be trivially satisfied.

Returning briefly to Dempster's rule of combination, as indicated above this rule is intended to allow the combination of 'independent' evidences and indeed plays a central role in the theory of DS-belief. Unfortunately the version of 'independence' required to make this rule compatible with our earlier justification of DS-belief functions is still a matter of contention, even if we generalise our justification to allow the possibility that the message '$\theta$ is true' might be incorrect and now take $Bel(\theta)$ to be the probability of the agent being led to conclude, on the basis of the message, that $\theta$ is true.

In view of the representation given in theorem 4.3 of a DS-belief function $Bel$ as the set of probability functions $W(Bel)$, it would seem natural to define the derived conditional DS-belief, $Bel(\theta||\phi)$ (so denoted to distinguish it from the previous function), by

$$Bel(\theta||\phi) = \inf\{w(\theta|\phi)|\ w \in W(Bel)\}.$$

Notice that provided $Bel(\phi) > 0$ this is well defined since $w(\phi) \geq Bel(\phi)$ for $w \in W$. What is much less obvious is that if $Bel(\phi) > 0$ then $Bel(\theta||\phi)$, as a function of $\theta \in SL$, is also DS-belief function. This surprising result is due to Fagin and Halpern [14] and (independently) Jaffray, see [14].

Before we give a proof of this, notice that if $w \in W(Bel)$ and $Bel(\phi) > 0$ then

$$Bel(\theta \wedge \phi) \leq w(\theta \wedge \phi),$$

$$w(\neg\theta \wedge \phi) \leq Pl(\neg\theta \wedge \phi),$$

by theorem 4.3, and

$$Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi) > 0$$

since if $m$ is the bpa of $Bel$ then

$$0 < Bel(\phi) = \sum_{\overline{\psi} \leq \overline{\phi}} m(\overline{\psi}) = \sum_{\overline{\psi} \leq \overline{\theta} \wedge \overline{\phi}} m(\overline{\psi}) + \sum_{\substack{\overline{\psi} \leq \overline{\phi} \\ \mathbf{0} < \overline{\psi} \wedge \overline{\neg\theta}}} m(\overline{\psi})$$

$$\leq \sum_{\overline{\psi} \leq \overline{\theta} \wedge \overline{\phi}} m(\overline{\psi}) + \sum_{\mathbf{0} < \overline{\psi} \wedge \overline{\neg\theta} \wedge \overline{\phi}} m(\overline{\psi}) = Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi).$$

Hence

$$\begin{aligned} w(\theta|\phi) &= \frac{w(\theta \wedge \phi)}{w(\theta \wedge \phi) + w(\neg\theta \wedge \phi)} \\ &\geq \frac{w(\theta \wedge \phi)}{w(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi)} \\ &\geq \frac{Bel(\theta \wedge \phi)}{Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi)}. \end{aligned}$$

Furthermore this lower bound is attained since by adapting the proof of theorem 4.3(ii) so that $\beta_{\overline{\psi}} \in S_{\psi \wedge \neg\theta \wedge \phi}$ if possible, otherwise $\beta_{\overline{\psi}} \in S_{\psi \wedge (\neg\theta \vee \neg\phi)}$ if possible, we can produce $w \in W(Bel)$ such that

$$w(\theta \vee \neg\phi) = Bel(\theta \vee \neg\phi),$$

$$w(\theta \wedge \phi) = Bel(\theta \wedge \phi).$$

Hence

$$w(\theta|\phi) = \frac{w(\theta \wedge \phi)}{w(\theta \wedge \phi) + 1 - w(\theta \vee \neg\phi)} = \frac{Bel(\theta \wedge \phi)}{Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi)}$$

as required.

It now follows that provided $Bel(\phi) > 0$,

$$Bel(\theta||\phi) = \frac{Bel(\theta \wedge \phi)}{Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi)} \tag{4.1}$$

**Theorem 4.4** *Let Bel be a DS-belief function on SL with $Bel(\phi) > 0$. Then $Bel(\theta||\phi)$, as a function of $\theta \in SL$, is a DS-belief function on SL.*

**Proof** $(DS1)$ and $(DS2)$ are straightforward to check. To show $(DS3)$ let $m$ be the bpa for $Bel$ and let

$$\{\overline{\theta}_1, \ldots, \overline{\theta}_t\} = \{\overline{\eta} \in \overline{SL} | m(\overline{\eta}) > 0, \overline{\eta} \leq \overline{\phi}\},$$

$$\{\overline{\phi}_1, \ldots, \overline{\phi}_s\} = \{\overline{\eta} \in \overline{SL} | m(\overline{\eta}) > 0, \overline{\eta} \not\leq \overline{\phi}, \mathbf{0} < \overline{\eta} \wedge \overline{\phi}\}.$$

Then by using (4.1) above,

$$Bel(\theta||\phi) = \frac{\sum_{\overline{\theta}_j \leq \overline{\theta} \wedge \overline{\phi}} m(\overline{\theta}_j)}{\sum_{\overline{\theta}_j \leq \overline{\theta} \wedge \overline{\phi}} m(\overline{\theta}_j) + \sum_{\mathbf{0} < \overline{\theta}_j \wedge \neg\overline{\theta} \wedge \overline{\phi}} m(\overline{\theta}_j) + \sum_{\mathbf{0} < \overline{\phi}_j \wedge \neg\overline{\theta} \wedge \overline{\phi}} m(\overline{\phi}_j)}$$

$$= \frac{\sum_{\overline{\theta}_j \leq \overline{\theta} \wedge \overline{\phi}} m(\overline{\theta}_j)}{D - \sum_{\overline{\phi}_j \wedge \overline{\phi} \leq \overline{\theta}} m(\overline{\phi}_j)}$$

where $D = \sum_{j=1}^{t} m(\overline{\theta}_j) + \sum_{j=1}^{s} m(\overline{\phi}_j)$.

Hence to show $(DS3)$, i.e.

$$Bel\left(\bigvee_{i=1}^{m} \psi_i \middle\| \phi\right) \geq \sum_{\emptyset \neq I \subseteq \{\psi_1, \ldots, \psi_m\}} (-1)^{|I|-1} Bel\left(\bigwedge I \middle\| \phi\right),$$

we must show that

$$\frac{\sum_{\overline{\theta}_j \leq \overline{\phi} \wedge \overline{\bigvee \psi_i}} m(\overline{\theta}_j)}{D - \sum_{\overline{\phi}_j \wedge \overline{\phi} \leq \overline{\bigvee \psi_i}} m(\overline{\phi}_j)} \geq \sum_{\emptyset \neq I \subseteq \{\psi_1, \ldots, \psi_m\}} (-1)^{|I|-1} \frac{\sum_{\overline{\theta}_j \leq \overline{\phi} \wedge \overline{\bigwedge I}} m(\overline{\theta}_j)}{D - \sum_{\overline{\phi}_j \wedge \overline{\phi} \leq \overline{\bigwedge I}} m(\overline{\phi}_j)}. \tag{4.2}$$

Let $\overline{\theta}_j \leq \overline{\phi} \wedge \bigvee \overline{\psi}_i$ and, without loss of generality, suppose that $\psi_1, \ldots, \psi_k$ are those $\psi_i$ such that $\overline{\theta}_j \leq \overline{\phi} \wedge \overline{\psi}_i$. Then by isolating the terms in (4.2) in which $m(\overline{\theta}_j)$ occurs it is clearly enough to show that

$$\frac{1}{D - \sum_{\overline{\phi}_j \wedge \overline{\phi} \leq \overline{\bigvee \psi_i}} m(\overline{\phi}_j)} \geq \sum_{\emptyset \neq I \subseteq R} (-1)^{|I|-1} \frac{1}{D - \sum_{\overline{\phi}_j \wedge \overline{\phi} \leq \overline{\bigwedge I}} m(\overline{\phi}_j)} \tag{4.3}$$

where $R = \{\psi_1, \ldots, \psi_k\}$.

In order to show (4.3) we now appeal to a rather technical lemma.

**Lemma 4.5** *Let $S, C_1, \ldots, C_k$ be sets of strictly positive reals. Then*

$$\sum_{I \subseteq \{C_1, \ldots, C_k\}} (-1)^{|I|}(x - \sum\{y \in S | y \in \bigcap I\})^{-m} \geq 0$$

*whenever $m > 0$ and $x > \sum_{y \in S} y$.*

**Proof.** The proof is by induction on $|S|$ (for all $C_1, \ldots, C_k$ and $m > 0$). For $|S| = 0$, i.e. $S = \emptyset$, the expression becomes

$$\sum_{I \subseteq \{C_1, \ldots, C_k\}} (-1)^{|I|} x^{-m}$$

which, as we have already seen, is $x^{-m}$ if $k = 0$ and zero otherwise. Either way it is non-negative, as required.

Now let $|S| = n > 0$ and assume the result for $n - 1$. Let $a \in S$ and without loss of generality let $C_1, \ldots, C_r$ be those $C_i$ such that $a \in C_i$. Then

$$\begin{aligned}
&\sum_{I \subseteq \{C_1, \ldots, C_k\}} (-1)^{|I|}(x - \sum\{y \in S | y \in \bigcap I\})^{-m} \\
&= \sum_{I \subseteq \{C_1, \ldots, C_r\}} \cdots + \sum_{\substack{I \subseteq \{C_1, \ldots, C_k\} \\ I \not\subseteq \{C_1, \ldots, C_r\}}} \cdots \\
&= \sum_{I \subseteq \{C_1, \ldots, C_r\}} (-1)^{|I|}((x - a) - \sum\{y \in S' | y \in \bigcap I\})^{-m} \\
&\quad + \sum_{\substack{I \subseteq \{C_1, \ldots, C_k\} \\ I \not\subseteq \{C_1, \ldots, C_r\}}} (-1)^{|I|}(x - \sum\{y \in S' | y \in \bigcap I\})^{-m}
\end{aligned}$$

where $S' = S - \{a\}$. This further equals

$$\sum_{I \subseteq \{C_1, \ldots, C_r\}} (-1)^{|I|}((x - a) - \sum\{y \in S' | y \in \bigcap I\})^{-m}$$

$$- \sum_{I \subseteq \{C_1,\dots,C_r\}} (-1)^{|I|}(x - \sum\{y \in S' | y \in \bigcap I\})^{-m}$$

$$+ \sum_{I \subseteq \{C_1,\dots,C_k\}} (-1)^{|I|}(x - \sum\{y \in S' | y \in \bigcap I\})^{-m}.$$

By inductive hypothesis the last of these three terms is non-negative. To show that the sum of the first two is also non-negative it suffices to show that

$$\sum_{I \subseteq \{C_1,\dots,C_r\}} (-1)^{|I|}(x - \sum\{y \in S' | y \in \bigcap I\})^{-m}$$

is decreasing (in $x$) for $x > \sum_{y \in S'} y$. But this follows since by using the inductive hypothesis again (for $m+1$) its derivative with respect to $x$ can be seen to be non-positive. □

Returning to the proof of (4.2) suppose for the moment that the numbers $m(\overline{\phi}_i)$ are all distinct. Then putting

$$S = \{m(\overline{\phi}_j) | \; \overline{\phi}_j \wedge \overline{\phi} \leq \bigvee_{i=1}^{m} \overline{\psi}_i\},$$

$$C_i = \{m(\overline{\phi}_j) | \; \overline{\phi}_j \wedge \overline{\phi} \leq \overline{\psi}_i\},$$

and $m = 1$, $x = D$ in the lemma and noticing that the left hand side of (4.3) can be written as

$$(-1)^{|I|}(D - \sum\{y \in S | y \in \bigcap I\})^{-1}$$

for $I = \emptyset$, we see that the theorem follows directly from the lemma. Clearly the result generalises to non-distinct $m(\overline{\phi}_i)$, for example by perturbing these values slightly and letting the perturbation go to zero. (The reason for not directly proving the required generalisation of the lemma in this case is the rather obfuscating notation which would need to be introduced.) □

It is easy to see that $Bel(\theta|\phi)$, $Bel(\theta||\phi)$ as defined above are different, even assuming $Bel(\phi) > 0$. For example if $L = \{p_1, p_2\}$, $\theta = p_1$, $\phi = p_2$ and the bpa $m$ of $Bel$ is given by $m(\overline{p_1}) = 1 - \delta$, $m(\overline{p_2}) = \delta > 0$ then $Bel(\phi) = \delta > 0$, $Bel(\theta||\phi) = 0$ but $Bel(\theta|\phi) = 1 - \delta$. For $Bel$ a probability function both agree with the standard conditional probability whilst in general $Bel(\theta|\phi) \geq Bel(\theta||\phi)$ and $Pl(\theta||\phi) \geq Pl(\theta|\phi)$. To see this notice, that by expressing both sides in terms of the bpa $m$ of $Bel$,

$$Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi) \leq Pl(\phi),$$

so, provided $Bel(\phi) > 0$,

$$\begin{aligned} Bel(\theta|\phi) &= 1 - Pl(\neg\theta|\phi) = 1 - \frac{Pl(\neg\theta \wedge \phi)}{Pl(\phi)} \\ &\geq 1 - \frac{Pl(\neg\theta \wedge \phi)}{Bel(\theta \wedge \phi) + Pl(\neg\theta \wedge \phi)} = Bel(\theta||\phi). \end{aligned}$$

Whilst the conditional belief $Bel(\theta||\phi)$ has a certain naturalness, it suffers from the (apparent) shortcoming that repeated conditionings do not commute. That is, even when well defined, conditioning with respect to $\phi_1$, $\phi_2$, in that order, is not necessarily the same as conditioning with respect to $\phi_2, \phi_1$, in that order, and hence not necessarily the same as the single conditioning on $\phi_1 \wedge \phi_2$. (This occurs for example in the case $L = \{p_1, p_2\}$, $\theta = p_1 \wedge p_2$, $\phi_1 = p_1$, $\phi_2 = p_1 \vee p_2$ with $Bel$ given by $m(p_1 \wedge p_2) = m(p_1 \wedge \neg p_2) = \frac{1}{4}, m(\neg p_1) = \frac{1}{2}$.) These three values are however always equal for $Bel(\theta|\phi)$, as can be readily checked.

For a further discussion on the relative suitability of $Bel(\theta|\phi)$, $Bel(\theta||\phi)$ for capturing the notion of conditional belief in the context of Dempster–Shafer belief see [11]. Notice that since the denominator in $Bel(\theta||\phi)$ involves $\theta$, linear constraints involving $Bel(\theta_1||\phi), \ldots, Bel(\theta_m||\phi)$ cannot in general be simplified to linear constraints in the unconditional belief function, unlike the other notions we have considered in this book.

We now turn to a second justification, or explanation, of DS-belief functions.

## Second Justification

In a similar fashion to the 'possible worlds' justification of belief as probability suppose we had a set $\mathcal{W}$ of possible worlds and a measure $\mu$ defined on a set $\mathcal{S}$ of subsets of $\mathcal{W}$ (rather than on all subsets of $\mathcal{W}$) with the properties:

(i) $\mathcal{W} \in \mathcal{S}$, $\emptyset \in \mathcal{S}$, $\mu(\mathcal{W}) = 1$, $\mu(\emptyset) = 0$.

(ii) If $X, Y \in \mathcal{S}$ then $X \cup Y$, $\mathcal{W} - X \in \mathcal{S}$ (so also $X \cap Y \in \mathcal{S}$).

(iii) If $X, Y \in \mathcal{S}$ and $X \cap Y = \emptyset$ then $\mu(X \cup Y) = \mu(X) + \mu(Y)$.

(That is, $\mu$ is a finitely additive measure on the field of sets $\mathcal{S}$.) Then for any $Z \subseteq \mathcal{W}$ (not necessarily in $\mathcal{S}$) we may define the *inner measure*, $\mu_*(Z)$, by

$$\mu_*(Z) = sup\{\mu(X) \mid X \in \mathcal{S} \text{ and } X \subseteq Z\}$$

and define, by analogy with the possible worlds interpretation of belief as probability,

$$Bel(\theta) = \mu_*\{w \in \mathcal{W} \mid w \models \theta\}.$$

The next two theorems, which appear in a paper, [17], of Fagin and Halpern, show that not only is $Bel$ defined in this way a DS-belief function but, further, that any DS-belief function has such a possible worlds interpretation.

**Theorem 4.6** *Bel defined as above is a DS-belief function.*

**Proof** We use the first definition of DS-belief functions. Conditions $(DS1\text{–}2)$ are immediate. To show $(DS3)$ let $\epsilon > 0$, $\theta_1, ..., \theta_n \in$ SL and for each $\overline{\psi} \in \overline{SL}$ pick $Z_{\overline{\psi}} \in \mathcal{S}$ such that

$$Z_{\overline{\psi}} \subseteq \{w \in \mathcal{W} \mid w \models \psi\}$$

and

$$\mu(Z_{\overline{\psi}}) \leq \mu_*\{w \in \mathcal{W} \mid w \models \psi\} \leq \mu(Z_{\overline{\psi}}) + \epsilon \tag{4.4}$$

We may assume that for $\overline{\phi} \leq \overline{\psi}$, $Z_{\overline{\phi}} \subseteq Z_{\overline{\psi}}$, otherwise replace $Z_{\overline{\psi}}$ by $\bigcup_{\overline{\phi} \leq \overline{\psi}} Z_{\overline{\phi}}$, this only tightens (4.4). Then

$$\begin{aligned} Bel(\bigvee_{i=1}^{n} \theta_i) &= \mu_*\{w \in \mathcal{W} \mid w \models \bigvee_{i=1}^{n} \theta_i\} \geq \mu(Z_{\overline{\theta}_1} \cup ... \cup Z_{\overline{\theta}_n}) \\ &= \sum_{\emptyset \neq S \subseteq \{1,...,n\}} (-1)^{|S|-1} \mu(\bigcap_{i \in S} Z_{\overline{\theta}_i}) \\ &\qquad \text{(this is proved as in proposition 2.1(d)))} \\ &\geq \sum_{\emptyset \neq S \subseteq \{1,...,n\}} (-1)^{|S|-1} \mu_*\{w \in \mathcal{W} \mid w \models \bigwedge_{i \in S} \theta_i\} - 2^n \epsilon \end{aligned}$$

since

$$Z_{\overline{\bigwedge_{i \in S} \theta_i}} \subseteq \bigcap_{i \in S} Z_{\overline{\theta}_i} \subseteq \{w \in \mathcal{W} \mid w \models \bigwedge_{i \in S} \theta_i\},$$

so

$$\mu\left(Z_{\overline{\bigwedge_{i \in S} \theta_i}}\right) \leq \mu\left(\bigcap_{i \in S} Z_{\overline{\theta}_i}\right) \leq \mu_*\{w \in \mathcal{W} \mid w \models \bigwedge_{i \in S} \theta_i\} \leq \mu(\bigcap_{i \in S} Z_{\overline{\theta}_i}) + \epsilon.$$

$(DS3)$ now follows since

$$\mu_*\{w \in \mathcal{W} \mid w \models \bigwedge_{i \in S} \theta_i\} = Bel(\bigwedge_{i \in S} \theta_i)$$

and $\epsilon > 0$ was arbitrary. □

An origin for $(DS3)$ is now clear; it is simply a reformulated version of an inequality which always holds for inner measures.

**Theorem 4.7** *If Bel is a DS-belief function then there are $\mathcal{W}, \mathcal{S}, \mu$ as above such that for all $\theta \in SL$,*

$$Bel(\theta) = \mu_*\{w \in \mathcal{W} \mid w \models \theta\}.$$

**Proof** Let $\mathcal{W} = \{\langle \alpha, Y\rangle \mid \alpha \in Y \subseteq At^L\}$ where, as usual, $At^L = \{\alpha_1, ..., \alpha_J\}$ is the set of atoms of $SL$. For $\emptyset \neq Y \subseteq At^L$ let

$$R_Y = \{\langle \alpha, Y\rangle \mid \alpha \in Y\}.$$

Notice that the $R_Y$ are pairwise disjoint, with union $\mathcal{W}$. Let $\mathcal{S}$ consist of all subsets of $\mathcal{W}$ of the form

$$R_{Y_1} \cup R_{Y_2} \cup ... \cup R_{Y_q} \quad (Y_i \text{ distinct})$$

and define $\mu(R_{Y_1} \cup R_{Y_2} \cup ... \cup R_{Y_q}) = \sum_{i=1}^{q} m(\overline{\bigvee Y_i})$, where $m$ is the bpa of $Bel$. It is straightforward to check that $\mathcal{S}$, $\mu$ satisfy (i)–(iii) above.

For $\langle \alpha, Y\rangle \in \mathcal{W}$ we define, for $\theta \in SL$,

$$\langle \alpha, Y\rangle \models \theta \quad \Leftrightarrow \quad \alpha \models \theta \text{ (in the usual sense)}.$$

Then

$$\begin{aligned}
\mu_*\{\langle \alpha, Y\rangle \in \mathcal{W} \mid \langle \alpha, Y\rangle \models \theta\} &= \mu_*\{\langle \alpha, Y\rangle \in \mathcal{W} \mid \alpha \models \theta\} \\
&= \mu_*\{\langle \alpha, Y\rangle \in \mathcal{W} \mid \alpha \in S_\theta\}.
\end{aligned}$$

Clearly $R_X \subseteq \{\langle \alpha, Y\rangle \in \mathcal{W} \mid \langle \alpha, Y\rangle \models \theta\}$ for $\emptyset \neq X \subseteq S_\theta$ and, conversely, if $R_X \subseteq \{\langle \alpha, Y\rangle \in \mathcal{W} \mid \langle \alpha, Y\rangle \models \theta\}$ then $\langle \beta, X\rangle \models \theta$ must hold for each $\beta \in X$ so $X \subseteq S_\theta$. Hence $\bigcup_{\emptyset \neq X \subseteq S_\theta} R_X$ is the largest set in $\mathcal{S}$ which is a subset of $\{\langle \alpha, Y\rangle \in \mathcal{W} \mid \langle \alpha, Y\rangle \models \theta\}$ so

$$\mu_*\{\langle \alpha, Y\rangle \in \mathcal{W} \mid \langle \alpha, Y\rangle \models \theta\} = \mu(\bigcup_{\emptyset \neq X \subseteq S_\theta} R_X) = \sum_{\emptyset \neq X \subseteq S_\theta} m(\overline{\bigvee X}) = Bel(\theta)$$

as required. □

**Remark** Notice that if $\mu_*\{w \in \mathcal{W} \mid w \models \theta\} = Bel(\theta)$ with $\mu$, etc. as above then

$$\begin{aligned}
Pl(\theta) = 1 - Bel(\neg\theta) &= 1 - \sup\{\mu(Z) \mid Z \in \mathcal{S} \text{ and } Z \subseteq X\} \\
&\quad\ \text{where } X = \{w \in \mathcal{W} \mid w \models \neg\theta\} \\
&= 1 - \sup\{\mu(\mathcal{W} - Z) \mid Z \in \mathcal{S} \text{ and } \mathcal{W} - Z \subseteq X\} \\
&= \inf\{1 - \mu(\mathcal{W} - Z) \mid Z \in \mathcal{S} \text{ and } \mathcal{W} - X \subseteq Z\} \\
&= \inf\{\mu(Z) \mid Z \in \mathcal{S} \text{ and } \{w \in \mathcal{W} \mid w \models \theta\} \subseteq Z\} \\
&= \mu^*\{w \in \mathcal{W} \mid w \models \theta\}
\end{aligned}$$

where $\mu^*$ is the *outer measure* on $\mathcal{W}$ derived from $\mu$.

**An Example** Recall the simple example of question $Q$ already considered in Chapter 2, namely, given

$$K = \{Bel(q) = b,\ Bel(p \mid q) = a\} \quad (b > 0)$$

what value should be assigned to $Bel(p)$? to $Bel(\neg p)$?

In this case if we further assume that $Bel$ is a DS-belief function, i.e. satisfies ($DS$1–3), and conditional belief is defined as above then the constraints become

$$\{Bel(q) = b,\ Bel(p \vee \neg q) - Bel(\neg q) = a(1 - Bel(\neg q))\}.$$

As indicated earlier we can re-express these in terms of the $m(\bar{\theta})$, $\bar{\theta} \in \overline{SL}$ and hence find the consistent ranges. However even for $L = \{p, q\}$ this involves $2^{2^2} - 1 = 15$ unknowns. Carrying out the calculation yields (maximal) ranges for $Bel(p), Bel(\neg p)$ of

$$1 + ab - b \geq Bel(p) \geq \max(0, a + b - 1),$$

$$1 - ab \geq Bel(\neg p) \geq 0.$$

A similar investigation using the conditional belief $Bel(p||q)$ yields ranges

$$1 + ab - b \geq Bel(p) \geq ab,$$

$$1 - ab \geq Bel(\neg p) \geq 0.$$