# A Defect in Dempster-Shafer Theory

**Pei Wang**
Center for Research on Concepts and Cognition, Indiana University
510 North Fess Street, Bloomington, IN 47408, USA
E-mail: *pwang@cogsci.indiana.edu*

**Abstract**

By analyzing the relationships among *chance*, *weight of evidence* and *degree of belief*, it is shown that the assertion "chances are special cases of belief functions" and the assertion "Dempster's rule can be used to combine belief functions based on distinct bodies of evidence" together lead to an inconsistency in Dempster-Shafer theory. To solve this problem, some fundamental postulates of the theory must be rejected. A new approach for uncertainty management is introduced, which shares many intuitive ideas with D-S theory, while avoiding this problem.

## 1 Introduction

Evidence theory, or Dempster-Shafer (D-S) theory, is developed as an attempt to generalize probability theory by introducing a rule for combining distinct bodies of evidence [1, 7].

The most influential version of the theory is presented by Shafer in his book *A Mathematical Theory of Evidence* [7]. In the book, the following postulates are assumed, which form the foundation of D-S theory.

**Postulate 1:** *Chance* is the limit of the proportion of "positive" outcomes among all outcomes [7, pages 9, 202].

**Postulate 2:** Chances, if known, should be used as *belief functions* [7, pages 16, 201].

**Postulate 3:** *Evidence combination* refers to the pooling, or accumulating, of distinct bodies of evidence [7, pages 8, 77].

**Postulate 4:** *Dempster's rule* can be used on belief functions for evidence combination [7, pages 6, 57].

In this paper, we show, by discussing a simple situation, that there is an inconsistency among the postulates. Then, we argue that though there are several possible solutions of this problem within the framework of D-S theory, each of them has serious disadvantages. Finally, we briefly introduce a new approach that achieves the goals of D-S theory, yet is still natural and consistent.

## 2 A Simplified Situation

In the following, we address only the simplest non-trivial *frame of discernment* $\Theta = \{H, H'\}$ ($|\Theta| = 1$ is trivial). Since $\Theta$ is exhaustive and exclusive by definition, we have $H' = \bar{H}$ (the negation of $H$).

In such a situation, all the information about the system's beliefs can be represented by a pair of real numbers on [0, 1]: the *degree of belief* and the *degree of plausibility* of $\{H\}$, $<Bel(\{H\}),\ Pl(\{H\})>$, and the belief about $\bar{H}$ can be derived from them. To simplify our notation, in the following the two numbers are referred to as $Bel$ and $Pl$.

$Bel$ and $Pl$ indicate the relationship between the hypothesis $H$ and the available evidence. When the system gets two *distinct bodies of evidence*, and they are measured by $<Bel_1,\ Pl_1>$ and $<Bel_2,\ Pl_2>$ respectively, then, after *evidence combination*, the pooled evidence is measured by the following $<Bel,\ Pl>$ value, according to Dempster's rule (**Postulate 4**):

$$\begin{aligned} Bel &= \frac{Bel_1 Pl_2 + Bel_2 Pl_1 - Bel_1 Bel_2}{1 - Bel_1(1 - Pl_2) - Bel_2(1 - Pl_1)} \\ Pl &= \frac{Pl_1 Pl_2}{1 - Bel_1(1 - Pl_2) - Bel_2(1 - Pl_1)} \end{aligned} \tag{1}$$

To specify the meaning of "evidence combination", Shafer introduces *weight of evidence*, $w$, with the following properties [7, pages 8, 88]:

1. $w$ is a measurement defined on bodies of evidence, with respect to a subset of $\Theta$, and it takes values on $[0,\ \infty]$.

2. When two entirely distinct bodies of evidence are combined, the weight of the pooled evidence (for the same subset of $\Theta$) is the *sum* of the original ones.

Therefore, if we use $w^+$ and $w^-$ to indicate the weight of evidence for $\{H\}$ and $\{\bar{H}\}$, respectively, then in the current situation the combination rule must satisfy the following relation (the **Postulate 3**):

$$\begin{aligned} w^+ &= w_1^+ + w_2^+ \\ w^- &= w_1^- + w_2^- \end{aligned} \tag{2}$$

where the subscripts 1 and 2 indicate bodies of evidence before the combination, as in (1).

The intuition behind the introducing of *weight of evidence* and **Postulate 3** is clear: we cannot apply an arbitrary rule for *evidence combination*, unless it captures the common usage of the notion, that is, by *combination*, the evidence is *pooled* or *accumulated*. Mathematically speaking, a certain measurement of the evidence (call it *weight*) is *additive* during the process.

Of course, the rule cannot be applied anywhere. We need to make sure that no evidence is repeatedly counted. This is what Dempster calls "independent sources of information" [1] and Shafer calls "distinct body of evidence" [7].

According to D-S theory, belief functions are determined by available evidence. Given (1) and (2), Shafer shows that the relationship between $<Bel,\ Pl>$ and $<w^+,\ w^->$ can be derived [7, page 84]:

$$\begin{aligned} Bel &= \frac{e^{w^+} - 1}{e^{w^+} + e^{w^-} - 1} \\ Pl &= \frac{e^{w^+}}{e^{w^+} + e^{w^-} - 1} \end{aligned} \tag{3}$$

It is also possible to derive (1) from (2) and (3), or derive (2) from (1) and (3). Therefore, the notion of evidence combination, the combination rule, and the relationship between weight of evidence and degree of belief are mutually determined.

Generally, we have $Bel \leq Pl$, or identically, $Bel(\{H\}) + Bel(\{\bar{H}\}) \leq 1$. When $Bel = Pl$, $Bel(\{H\})$ become a probability function, because then $Bel(\{H\}) + Bel(\{\bar{H}\}) = 1$. In [1], Dempster calls such a belief function "sharp," and treats it as "an ordinary probability measure." In [7], Shafer calls it "Bayesian," and writes it as $Bel_\infty(\{H\})$.

From (3), it is clear that $Bel = Pl$ happens if and only if the weight of all evidence, $w$ ($w = w^+ + w^-$), goes to infinite:

$$Bel_\infty(\{H\}) = \lim_{w \to \infty} Bel = \lim_{w \to \infty} Pl \tag{4}$$

Shafer interprets the above relationship as indicating that probability functions is a subset of belief functions [7, pages 19], and degree of belief converges to chance when the available evidence goes to infinite (**Postulate 2**), that is,

$$Bel_\infty(\{H\}) = Pr(H) \tag{5}$$

where $Pr(H)$ is the *chance*, or *aleatory probability*, of $H$ [7, pages 16, 33, 201].

In D-S theory, "chance" is used with its usual meaning as in statistics: for an experiment, if $t$ is the number of outcomes, and $t^+$ is the number of outcomes that correspond to $H$, then (**Postulate 1**)

$$Pr(H) = \lim_{t \to \infty} \frac{t^+}{t} \tag{6}$$

# 3 A Problem

From the above descriptions, D-S theory seems to be a reasonable extension of probability theory, because it introduces a combination rule, and still converges to probability theory when $Bel(\{H\})$ and $Pl(\{H\})$ overlap.

To see clearly how D-S theory and probability theory are related to each other, consider the situation where evidence for $H$ is in the form of a sequence of experiment outcomes with the following properties:

1. No single outcome can completely confirm or refute $H$.
2. There are only two possible outcomes: one supports $H$, while the other supports $\bar{H}$.
3. The outcomes provide distinct bodies of evidence.

Because there are only two types of evidence, we can assign two positive real numbers $w_0^+$ and $w_0^-$ as weights of evidence to an outcome supporting $H$ and $\bar{H}$, respectively. After $t$ outcomes are observed, in which $t^+$ outcomes support $H$ and $t^-$ outcomes support $\bar{H}$ ($t^+ + t^- = t$), the weight of available positive, negative and total evidence (for $H$) can be calculated according to **Postulate 3**:

$$\begin{aligned} w^+ &= w_0^+ t^+, \\ w^- &= w_0^- t^-, \\ w &= w^+ + w^-. \end{aligned}$$

When $t$ goes to infinity so does $w$, and vice versa. If $t^+/t$ converges to a limit $Pr$, then according to **Postulate 1** and **Postulate 2**, $Bel$ and $Pl$ should also converge to $Pr$, to become $Bel_\infty(\{H\})$.

We can rewrite $w^+$ and $w^-$ as functions of $t$ and $t^+$ in the relationships between belief function and weight of evidence (3), which is derived from **Postulate 3** and **Postulate 4**. If we then take the limit of the equation when $t$ (as well as $w$) goes to infinity, we get

$$\begin{aligned} Bel_\infty(\{H\}) &= \lim_{w \to \infty} \frac{e^{w^+} - 1}{e^{w^+} + e^{w^-} - 1} \\ &= \lim_{t \to \infty} \frac{e^{w_0^+ t^+} - 1}{e^{w_0^+ t^+} + e^{w_0^- t^-} - 1} \\ &= \begin{cases} 0 & \text{if } w_0^+ Pr < w_0^-(1 - Pr) \\ 1 & \text{if } w_0^+ Pr > w_0^-(1 - Pr) \\ \frac{1}{1+e^\Delta} & \text{if } w_0^+ Pr = w_0^-(1 - Pr) \end{cases} \end{aligned} \tag{7}$$

where $\Delta = \lim_{t \to \infty}(w_0^- t^- - w_0^+ t^+)$. The appendix contains the details in the last step.

This means that if $Pr$ (the chance of $H$ defined by **Postulate 1**) exists, then, by repeatedly applying Dempster's rule to combine the coming evidence (provided by the outcomes of the experiment), both $Bel$ and $Pl$ will converge to a point only when $\lim_{t \to \infty}(w_0^- t^- - w_0^+ t^+)$ exists, and even in that case $Bel_\infty(\{H\})$ is not $Pr$ in most cases, but 0, 1, or $1/(1 + e^\Delta)$, indicating qualitatively whether there is more positive evidence than negative evidence. The conclusion $Bel_\infty = 1/(1 + e^\Delta)$ is proven by Shafer himself [7, page 198]. However, he does not relate it to chance.

Therefore, in contrary with the **Postulate 2**, $Bel_\infty(\{H\})$ is usually different from $Pr(H)$, unless $Pr(H)$ happens to be 0, 1, or $w_0^-/(w_0^+ + w_0^-)$, and in the last case $w_0^- t^- - w_0^+ t^+$ must have a limit $\Delta$ that makes $1/(1 + e^\Delta) = Pr(H)$.

Therefore, though a Bayesian belief function is indeed a probability function in the sense that $Bel(\{H\}) + Bel(\{\bar{H}\}) = 1$, it is usually different from the chance of $H$.

This inconsistency is derived from the four postulates alone, so it is independent from other controversial issues, such as the interpretation of belief function, the accurate definition of "distinct" bodies of evidence, and the actual measurement of weight of evidence. No matter what opinions are accepted on these issues, as long as they are held consistently, the previous problem remains. For example, the choice of $w_0^+$ and $w_0^-$ can only determine which chance value is mapped to the degree of belief $1/(1+e^\Delta)$ (so all the other values are mapped to 0 or 1 correspondingly), but cannot change the result that chance and Bayesian belief function are usually different.

A possible argument against the above demonstration is to interpret "distinct bodies of evidence" in such a way that it is invalid to apply Dempster's rule in the previous situation. For example, according to Smets, "distinctness" is not satisfied in the present context because of the existence of a underlying probability function $Pr$ that create a link among the outcomes of the experiment [13]. Accepting such an opinion, however, means that **Postulate 2** is rejected. How can we say that "chances are limits of belief functions," if it is *always invalid* to take this kind of limits (by repeatedly applying Dempster's rule on the belief functions)?

The discrepancy also unearths some other inconsistencies in D-S theory. For example, Shafer describes *chance* as "essentially hypothetical rather than empirical," and unreachable by collecting (finite) evidence [7, page 202]. According to this interpretation, combining the evidence of two different Bayesian belief functions becomes invalid or nonsense, because they are chances and therefore not supported by finite empirical evidence. If $Bel_{\infty 1}(\{H\})$ and $Bel_{\infty 2}(\{H\})$ are different, then they are two conflicting conventions, and applying Dempster's rule to them is unjustified. If $Bel_{\infty 1}(\{H\})$ and $Bel_{\infty 2}(\{H\})$ are equal, then they are the same convention made from different considerations. In D-S theory, however, they are combined to get a different Bayesian belief function, except for some special points. Such a result is counter-intuitive [18] and inconsistent with Shafer's interpretation of chance.

There are already many papers on the justification of Dempster's rule [2, 5, 6, 11, 14, 19], but few of them addresses the relationships among degree of belief, weight of evidence, and chance. As a result, the mathematical properties of D-S theory are explored in detail, but its usage of notions, such as "chance" and "evidence combination", lacks a careful analysis. For instance, **Postulate 2** is usually accepted because a Bayesian belief function is indeed *a probability function*, but it is ignored that it is usually not equal to *the chance*, defined by **Postulate 1**.

# 4 Possible Solutions

It is always possible to save a theory, if we do not mind to twist or redefine the involved concepts. To solve the current problem, at least one of the four postulates in D-S theory must be removed. In the following, let us check all four logical possibilities one by one.

It seems unpopular to reject **Postulate 1**, and redefine "chance" as $\lim_{w\to\infty}(e^{w^+} - 1)/(e^{w^+} + e^{w^-} - 1)$, though this will lead to a consistent theory. The reason is simple: to use "chance" for the limit of the proportion of positive evidence is a well accepted convention, and a different usage of the concept will cause many confusions.

How about **Postulate 3**? In the following, we can see that if the *addition* of weight of evidence, during the combination of evidence from distinct sources, is replaced by *multiplication*, we can also get a consistent theory.

Let us assume $w^+ = w_1^+ w_2^+$ and $w^- = w_1^- w_2^-$ when two Bayesian belief functions $Bel_{\infty 1}$ and $Bel_{\infty 2}$ are combined to become $Bel_\infty$. Now, if we simply use the number of outcomes as weight of evidence, then from **Postulate 1**, **Postulate 2**, and the new postulate, we get

$$\begin{aligned} Bel_\infty &= \lim_{w \to \infty} \frac{w_1^+ w_2^+}{w_1^+ w_2^+ + w_1^- w_2^-} \\ &= \frac{Bel_{\infty 1} Bel_{\infty 2}}{Bel_{\infty 1} Bel_{\infty 2} + (1 - Bel_{\infty 1})(1 - Bel_{\infty 2})}. \end{aligned}$$

which is a special case of (1) (**Postulate 4**), when $Bel_{\infty 1} = Bel_1 = Pl_1$ and $Bel_{\infty 2} = Bel_2 = Pl_2$ (for Bayesian belief functions).

Though we preserve consistency, the result is not intuitively appealing. For example, no matter how the weight of evidence is actually measured, the combination of two pieces of positive evidence with unit weight ($w_1^+ = w_2^+ = 1$) will get $w^+ = 1$. That is, evidence is no longer accumulated by combination ($w^+$ may even be less than $w_1^+$, if $w_2^+ < 1$). This is not what we have in mind when talking about evidence combination or pooling.

Another way to reject **Postulate 3** is to remove the concept of *weight of evidence* from D-S theory. Actually weight of evidence is seldom mentioned in the literature of D-S theory. Shafer, in his later papers (for example, [9, 10]), tends to relate belief functions to *reliability of testimony* and *randomly coded message*, rather than to *weight of evidence*. One problem of such a solution is the loss of the intuition in the notion of "evidence combination". As discussed before, by "combination", we usually mean "pooling", "accumulating", or "putting together", and to introduce a measurement on evidence, which remains additive during combination, is important for justifying that the "combination rule" is really carrying out an operation on belief functions that corresponds to what we mean by "evidence combination" in everyday language. Without such a measurement, the claim that a rule does "evidence combination" is much less convincing. On the other hand, without weight of evidence, the problem is still there. In the previous example, if we directly assign $Bel$ and $Pl$ values to the two types of outcomes (rather than assign weights of evidence to them, as we do in the previous section), then use Dempster's rule to combine the belief functions, it can be proven, in a similar way as (7) is proven, that $Bel$ and $Pl$ usually do not converge to $Pr$. This is the case because of the one-to-one correspondence between the weight of evidence and the belief function.

The rejection of **Postulate 2** seems more plausible than the previous alternatives. Very few authors actually use $Bel_\infty(\{H\})$ to represent the chance of $H$. Even in Shafer's classic book [7], in which **Postulate 2** is made or assumed at several places, $Bel_\infty$ is not directly applied to represent statistical evidence.

However, there is not a consensus in the "Uncertainty in AI" community that $Bel_\infty(\{x\})$ and $Pr(x)$ are unequal. The following phenomena shows this:

1. According to many, if not all, textbooks and introductory papers, D-S theory is a generalization of probability theory, and a chance can be used as a degree of belief.

2. The "lower-upper bounds of probability" interpretation for belief functions is still accepted by some authors [4].

3. Some other authors, including Shafer himself, reject the above interpretation, but they still refer to a probability function as a special type (or a limit) of belief functions [9].

4. Though some authors have gone so far to the conclusion that Bayesian belief functions do not generally correspond to Bayesian measures of belief, they still view a belief function as the lower bound of probability [18].

5. In the transferable belief model of D-S theory [11, 12, 13], Smets shows that it is possible "for quantified beliefs developed independently of any underlying probabilistic model," though he still believes that "it seems reasonable to defend the idea that the belief of an event should be numerically equal to the probability of that event" [13].

Although it is possible to get rid of the inconsistency by give up the equality of $Bel_\infty(\{x\})$ and $Pr(x)$, such a solution will make the relationship between probability theory and D-S theory complicated.

If we accept **Postulate 1**, **Postulate 3**, **Postulate 4**, and the assumption that $w_0^+ = w_0^- = 1$ (this is assumed only to simplify the derivation), then from (3), the proportion of positive evidence of $H$ can be represented as a function of $Bel$ and $Pl$, when $Bel < Pl$, as

$$\frac{w^+}{w} = \frac{\log Pl - \log(Pl - Bel)}{\log Pl + \log(1 - Bel) - 2\log(Pl - Bel)}$$

Still, the relationship is not natural, and the ratio usually does not converge to the same point with $Bel$ and $Pl$ as evidence comes. As a result, a natural way to represent uncertainty as *proportion of positive evidence* becomes less available in D-S theory. As shown before, $Bel(\{H\})$ is more sensitive to the difference of $w^+$ and $w^-$, than to the proportion $w^+/w$. $Pr(H)$, as the limit of the proportion, even cannot be represented. The knowledge "$Pr(H) = 0.51$" and "$Pr(H) = 0.99$" will both be represented as $Bel(\{H\}) = Pl(\{H\}) = 1$, and their difference will be lost.

If **Postulate 2** were rejected, it would be invalid to interpret $Bel$ and $Pl$ as "lower and upper probability" [1, 3, 4, 12]. It is true that there are probability functions $P(x)$ satisfying

$$Bel(\{x\}) \leq P(x) \leq Pl(\{x\}), \text{ for all } x \in \Theta.$$

However, as demonstrated above, these functions may be unrelated to $Pr(H)$.

For the same reason, the assertion that "the Bayesian theory is a limiting case of D-S theory" [7, page 32] may be misleading. From a mathematical point of view, this assertion is true, since $Bel_\infty(\{H\})$ is a probability function. But as discussed previously, it is not the probability, or chance, of $H$. Therefore, it is not valid to get inference rules for D-S theory by extending Bayes theorem. In general, the relationship between D-S theory and probability theory will be very loose.

It is still possible to put different possible probability distributions into $\Theta$ and to assign belief function to them, as Shafer did [7, 8]. For example, the knowledge "$Pr(H) = 0.51$" can be represented as "$Bel(\{Pr(H) = 0.51\}) = 1$." However, here the probability function is *evaluated* by the belief function, rather than being a special case of it. The two are at different levels. As a result, the initial idea of D-S theory (to generalize probability theory), no longer holds. From a practical point of view, this approach is not appealing, neither. For instance, for any evidence combination to occur there must be *finite* possible probabilities for $H$ at the very beginning. It is unclear how to get them.

Finally, it is unlikely, though not completely impossible, to save D-S theory by rejecting **Postulate 4**. For instance, we can say that Dempster's rule does not apply to evidence combination, but can be used for some other purposes. Even so, the initial goal of D-S theory will be missed. Another suggestion is to use Dempster's rule only on non-Bayesian belief functions [13, 18]. However, the problem remains under the constraint, because in the previous demonstration Dempster's rule is only applied to non-Bayesian belief functions to make equations (3) true.

In summary, though it is possible for D-S theory to survive the inconsistency by removing one of the postulates, the result is still unsatisfactory. Either the natural meaning of "chance" or "evidence combination" must be changed, or the theory will fail to meet its original purpose, that is, to extend probability theory by introducing an evidence combination rule.

# 5 An Alternative Approach

In spite of the problems, some intuitions behind D-S theory are still attractive, such as the first three postulates, the idea of lower-upper probabilities [1], and the distinction between disbelief and lack of belief [7].

From previous discussion, we have seen that the core of evidence combination is the relationships among degree of belief, chance, and weight of evidence. The combination rule can be derived from these relationships.

Let us continue with the previous example. Because all the measurements are about $H$, we will omit it to simplify the formulas. Following the practice of statistics, for the current example a very natural convention is to use the number of outcomes as the weight of evidence, that is, to let $w_0^+ = w_0^- = 1$.

Because our belief about $H$ is totally determined by available evidence, it may be uncertain due to the existence of negative evidence. To measure the relative support that $H$ gets from available evidence, the most often used method is to take the *frequency* of positive evidence: $f = w^+/w$. According to **Postulate 1**, $\lim_{w \to \infty} f = Pr$, that is, the limit of $f$, if it exists, is the probability, or chance, of $H$. Therefore, we can refer to frequency as probability generalized to the situation of finite evidence.

However, when evidence combination is considered, $f$ alone cannot capture the uncertainty about $H$. When new evidence is combined with previous evidence, $f$ must be reevaluated. If we only know its previous value, we cannot determine how much it should be changed — the absolute amount of evidence is absent in $f$. Though it is possible, in theory, to directly use $w$ and $w^+$ as measurements of uncertainty, it is often unnatural and inconvenient [17]. Can we capture this kind of information without recording $w$ and $w^+$ directly?

Yes, we can. From the viewpoint of evidence combination, the influence of $w$ appears in the *stability* of a frequence evaluation based on it. Let us compare two situations: in the first $w = 1000$ and $w^+ = 600$, and in the second $w = 10$ and $w^+ = 6$. Though in both cases $f$ is 0.6, its stability is quite different. After a new outcome is observed, in the first situation the new frequency becomes either 600/1001 or 601/1001, while in the second it is either 6/11 or 7/11. The adjustment is much larger in the second situation than in the first.

If the information about stability is necessary for evidence combination, why not directly use intervals like [600/1001, 601/1001] and [6/11, 7/11] to represent the uncertainty in the previous situations?

Generally, let us introduce a pair of new measurements: a *lower frequency*, $l$, and a *upper frequency*, $u$, which are defined as

$$\begin{aligned} l &= \frac{w^+}{w+1} \\ u &= \frac{w^+ + 1}{w+1}. \end{aligned} \tag{8}$$

The idea behind $l$ and $u$ is simple: if the current frequency is $w^+/w$, then, after combining the current evidence (whose weight is $w$) with the new evidence provided by a new outcome (whose weight is 1), the new frequency will be in the interval $[l, u]$. We use an interval instead of a pair of points because the measurements will be extended to situations in which the weights of evidence are not necessarily integers. In general, the interval bounds the frequence until the weight of new evidence reaches a constant unit. For the current purpose, the 1 that appears in the definitions of $l$ and $u$ can be substituted by any positive number [17]. 1 is used here to simplify the discussion.

As bounds of frequency, $l$ and $u$ share intuitions with Dempster's $P_*$ and $P^*$, as well as Shafer's $Bel$ and $Pl$. However, they have some properties that distance them from the functions of D-S theory and other similar ideas like lower and upper bounds of probability:

1. $l \leq f \leq u$, that is, the current frequency is within the $[l, u]$ interval. Furthermore, it is easy to see that $f = l/(1-u+l)$, so the frequency value can be easily retrieved from the bounds.

2. The bounds of frequency are defined in terms of available evidence, which is finite. Whether the frequence of positive evidence really has a limit does not matter. The interval is determined before the next outcome occurs.

3. $\lim_{w\to\infty} l = \lim_{w\to\infty} f = \lim_{w\to\infty} u = Pr$. If $f$ does have a limit $Pr$, then $Pr$ is also the limit of $l$ and $u$. Therefore, probability is a special case of the $[l, u]$ interval, in which the interval degenerates into a point.

4. However, $Pr$, if it exists, is not necessarily in the interval all the time that evidence is accumulating. $[l, u]$ indicates the range $f$ will be from the current time to a *near future* (until the weight of new evidence reaches a constant), not an *infinite future*. Therefore, $l$ and $u$ are not bounds of probability.

5. The width of the interval $i = u - l = 1/(w+1)$ monotonically decreases during the accumulating of evidence, and so can be used to represent the system's "degree of ignorance" (about $f$). When $w = 0$, $i = 1$, because with no evidence, ignorance reaches its maximum. When $w \to \infty$, $i = 0$, because with infinite evidence the probability is obtained, so the ignorance (about the frequency) reaches its minimum, even though the next outcome is still uncertain. In this way, "lack of belief" and "disbelief" are clearly distinguished.

From the definitions of the lower-upper frequencies and **Postulate 3**, a combination rule, from $[l_1, u_1] \times [l_2, u_2]$ to $[l, u]$, is uniquely determined in terms of lower-upper frequencies, when neither $i_1 = u_1 - l_1$ nor $i_2 = u_2 - l_2$ is 0:

$$\begin{aligned} l &= \frac{l_1 i_2 + l_2 i_1}{i_1 + i_2 - i_1 i_2} \\ u &= \frac{l_1 i_2 + l_2 i_1 + i_1 i_2}{i_1 + i_2 - i_1 i_2}. \end{aligned} \tag{9}$$

From (3) and (8), we can even set up a one-to-one mapping between the *Bel*-*Pl* scale and the $l$-$u$ scale, when the weight of evidence $w$ is finite and $|\Theta| = 2$. In this way, the combination rule given by (9) is mapped exactly onto Dempster's rule (1). From a mathematical point of view, the two approaches differ only when $w \to \infty$. Then $Bel$ and $Pl$ converge to a probability if and only if $w^- - w^+$ converges to a constant, but $l$ and $u$ converge to a probability if and only if $w^+/w$ converges to a constant. The latter, being the probability of $H$, is more helpful and important in most situations than the former is. In fact, Shafer acknowledges the problem when he writes, "It is difficult to imagine a belief function such as $Bel_\infty$ being useful for the representation of actual evidence [7, page 199]." However, the result seems to be accepted without further analysis, since it follows from Dempster's rule.

Let us apply the paradigm to infinite evidence. For practical purpose it is impossible for a system to get infinite evidence, but we can use this concept to put definitions and conventions into a system. Beliefs supported by infinite evidence can be processed as normal ones, but will not be changed through evidence combinations.

According to the interpretation of the $[l,\ u]$ interval, it is not difficult to extend the new combination rule (9) to the case of infinite evidence:

1. When $i_1 = 0$ but $i_2 > 0$, the rule is still applicable in the form of (9), which gives the result that $l = l_1 = u_1 = u$. Thus when uncertainty is represented by probability (a point, instead of an interval), it will not be effected by combining its evidence with finite new evidence.

2. When $i_1 = i_2 = 0$, the rule cannot be used. Now the system will distinguish two cases:

   (a) when $l_1 = l_2 = u_1 = u_2$ there are two identical probabilistic judgments, so one of them can be removed (because it is redundant), leaving the other as the conclusion; or,

   (b) $l_1 \neq l_2$, meaning there are two conflicting probabilistic judgments. Since such judgments are not generated from evidence collection but from conventions or definitions, the two judgments are not "combined," but reported to the human/program which is responsible for making the conventions.

Here we are even more faithful to Shafer's interpretation of (aleatory) probability than D-S theory is. Being "essentially hypothetical rather than empirical," probability cannot be evaluated with less than infinite evidence [7, page 201]. For the same reason, it should not be changed by less than infinite evidence.

In summary, though many of the intuitive ideas of D-S theory are preserved, the problem in D-S theory discussed above no longer exists in the "lower-upper frequency" approach. The new method can represent probability and ignorance, and has a rule for evidence combination. The new approach can hardly be referred to as a modification or extension of D-S theory, in part because Dempster's rule is not used.

This approach is used in the *Non-Axiomatic Reasoning System* (NARS) project. As an intelligent reasoning system, NARS can adapt to its environment and answer questions with insufficient knowledge and resources [16, 17]. A complete comparison of NARS and D-S theory is beyond the scope of this paper. By introducing the approach here, we hope to show that the most promising solution for the previous inconsistency is to reject **Postulate 4** and go beyond D-S theory.

# 6 Conclusion

A variety of authors have noticed that certain applications of D-S theory lead to counter-intuitive results. However, the origin of the problem is studied insufficiently.

Though D-S theory can be used to accumulate evidence from distinct sources, it establishes a unnatural relation between degree of belief and weight of evidence by using Dempster's rule for evidence combination. As a result, the assertion that "probability is a special belief function" is in conflict with the definitions of "probability" and "evidence combination."

The inconsistency is solvable within D-S theory, but such a solution will make D-S theory either lose its naturalness (by using a concept in a unusual way), or miss its original goals (by being unable to represent probability or to combine evidence).

Though the criticism of D-S theory to Bayes approach is justifiable, and the "lower-upper frequency" approach is motivated by similar theoretical considerations [15], the two approaches solve the problem differently.

The "lower-upper frequency" approach is not specially designed to replace D-S theory in general, but it does suggest a better way to represent and process uncertainty. The new approach sets up a more natural relation among the various measurements of uncertainty, including probability. It can combine evidence from distinct sources. Therefore, it makes the system capable of carrying out multiple types of inference, such as deduction, induction, and abduction [16, 17].

# APPENDIX: Detailed derivation of (7)

- If $w_0^+ Pr < w_0^-(1-Pr)$, then

$$\begin{aligned}\lim_{t\to\infty} \frac{e^{w_0^+ t^+} - 1}{e^{w_0^+ t^+} + e^{w_0^- t^-} - 1} &= \lim_{t\to\infty} \frac{e^{w_0^+ t^+ - w_0^- t^-} - e^{-w_0^- t^-}}{e^{w_0^+ t^+ - w_0^- t^-} + 1 - e^{-w_0^- t^-}} \\ &= \lim_{t\to\infty} \frac{e^{-t[w_0^-(1-\frac{t^+}{t}) - w_0^+ \frac{t^+}{t}]} - e^{-t[w_0^-(1-\frac{t^+}{t})]}}{e^{-t[w_0^-(1-\frac{t^+}{t}) - w_0^+ \frac{t^+}{t}]} + 1 - e^{-t[w_0^-(1-\frac{t^+}{t})]}} \\ &= \frac{0-0}{0+1-0} \\ &= 0\end{aligned}$$

- If $w_0^+ Pr > w_0^-(1-Pr)$, then

$$\begin{aligned}\lim_{t\to\infty} \frac{e^{w_0^+ t^+} - 1}{e^{w_0^+ t^+} + e^{w_0^- t^-} - 1} &= \lim_{t\to\infty} \frac{1 - e^{-w_0^+ t^+}}{1 + e^{w_0^- t^- - w_0^+ t^+} - e^{-w_0^+ t^+}} \\ &= \lim_{t\to\infty} \frac{1 - e^{-t[w_0^+ \frac{t^+}{t}]}}{1 + e^{-t[w_0^+ \frac{t^+}{t} - w_0^-(1-\frac{t^+}{t})]} - e^{-t[w_0^+ \frac{t^+}{t}]}} \\ &= \frac{1-0}{1+0-0} \\ &= 1\end{aligned}$$

- If $w_0^+ Pr = w_0^-(1-Pr)$, and $\lim_{t\to\infty}(w_0^- t^- - w_0^+ t^+) = \Delta$, then

$$\begin{aligned}\lim_{t\to\infty} \frac{e^{w_0^+ t^+} - 1}{e^{w_0^+ t^+} + e^{w_0^- t^-} - 1} &= \lim_{t\to\infty} \frac{1 - e^{-w_0^+ t^+}}{1 + e^{w_0^- t^- - w_0^+ t^+} - e^{-w_0^+ t^+}} \\ &= \frac{1}{1+e^{\Delta}}\end{aligned}$$

# Acknowledgment

This work is supported by a research assistantship from the Center for Research on Concepts and Cognition, Indiana University.

A previous version of the paper was presented in the *Tenth Conference on Uncertainty in Artificial Intelligence*, Seattle, Washington, July 1994. I have benefited from the discussions with the attendants of the conference, especially Benjamin Grosof, Judea Pearl, Philippe Smets, and Nic Wilson.

# References


[1] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[2] D. Dubois and H. Prade. Updating with belief functions, ordinal conditional functions and possibility measures. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 311–329. North-Holland, Amsterdam, 1991.

[3] D. Dubois and H. Prade. Evidence, knowledge, and belief functions. *International Journal of Approximate Reasoning*, 6:295–319, 1992.

[4] R. Fagin and J. Halpern. A new approach to updating beliefs. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 347–374. North-Holland, Amsterdam, 1991.

[5] H. Kyburg. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers, San Mateo, California, 1988.

[7] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.

[8] G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society. Series B*, 44:322–352, 1982.

[9] G. Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990.

[10] G. Shafer and A. Tversky. Languages and designs for probability judgment. *Cognitive Science*, 12:177–210, 1985.

[11] Ph. Smets. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:447–458, 1990.

[12] Ph. Smets. The transferable belief model and other interpretations of Dempster-Shafer's model. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 375–383. North-Holland, Amsterdam, 1991.

[13] Ph. Smets. Belief induced by the partial knowledge of the probabilities. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 523–530. Morgan Kaufmann Publishers, San Francisco, California, 1994.

[14] F. Voorbraak. On the justification of Dempster's rule of combination. *Artificial Intelligence*, 48:171–197, 1991.

[15] P. Wang. Belief revision in probability theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 519–526. Morgan Kaufmann Publishers, San Mateo, California, 1993.

[16] P. Wang. Non-axiomatic reasoning system (version 2.2). Technical Report 75, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993. Available via WWW at *http://www.cogsci.indiana.edu/farg/peiwang/papers.html*.

[17] P. Wang. From inheritance relation to nonaxiomatic logic. *International Journal of Approximate Reasoning*, 11(4):281–319, November 1994.

[18] N. Wilson. The combination of belief: when and how fast? *International Journal of Approximate Reasoning*, 6:377–388, 1992.

[19] N. Wilson. The assumptions behind Dempster's rule. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 527–534. Morgan Kaufmann Publishers, San Mateo, California, 1993.