Artificial Intelligence 205 (2013) 1–29

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Evidential reasoning rule for evidence combination

Jian-Bo Yang *, Dong-Ling Xu

Decision and Cognitive Sciences Research Centre, The University of Manchester, Manchester M15 6PB, UK

ARTICLE INFO

Article history:
Received 26 January 2013
Received in revised form 9 September 2013
Accepted 13 September 2013
Available online 23 September 2013

Keywords:
Evidential reasoning
Belief distribution
Dempster–Shafer theory
Bayesian inference
Multiple criteria decision analysis
Information fusion

ABSTRACT

This paper aims to establish a unique *Evidential Reasoning* (*ER*) rule to combine multiple pieces of independent evidence conjunctively with weights and reliabilities. The novel concept of *Weighted Belief Distribution* (*WBD*) is proposed and extended to *WBD* with *Reliability* (*WBDR*) to characterise evidence in complement of *Belief Distribution* (*BD*) introduced in *Dempster–Shafer* (*D–S*) theory of evidence. The implementation of the orthogonal sum operation on *WBD*s and *WBDR*s leads to the establishment of the new *ER* rule. The most important property of the new *ER* rule is that it constitutes a generic conjunctive probabilistic reasoning process, or a generalised Bayesian inference process. It is shown that the original *ER* algorithm is a special case of the *ER* rule when the reliability of evidence is equal to its weight and the weights of all pieces of evidence are normalised. It is proven that Dempster's rule is also a special case of the *ER* rule when each piece of evidence is fully reliable. The *ER* rule completes and enhances Dempster's rule by identifying how to combine pieces of fully reliable evidence that are highly or completely conflicting through a new reliability perturbation analysis. The main properties of the *ER* rule are explored to facilitate its applications. Several existing rules are discussed and compared with the *ER* rule. Numerical and simulation studies are conducted to show the features of the *ER* rule.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Providing an *ER* framework for evidence combination and information fusion in an *Artificial Intelligence* (*AI*) system has been an important task and attracted significant interests and efforts from communities in *AI*, computing, operational research, decision sciences, system sciences, control theory, information systems, etc. From all databases in the Web of Knowledge, the term *ER* first appeared in a paper published by *AI* in 1985 [15], although the term might well have been used in other context earlier. In their paper, Gordon and Shortliffe embraced the use of a *D–S* scheme for evidence-aggregation processes in a hypothesis space. In another paper published also in *AI* [24], it was shown that *ER* could be conducted in the same hypothesis space using a Bayesian scheme. However, the two schemes are different, and the nature and significance of their differences were investigated [20].

The *D–S* scheme is based on a frame of discernment composed of a set of propositions that are mutually exclusive and collectively exhaustive [25]. In the *D–S* scheme, basic probabilities can be assigned to not only singleton propositions but also any of their subsets, thereby allowing a piece of evidence to be profiled by a *BD* defined on the power set of the frame of discernment. *BD* is regarded as the most natural and flexible generalisation of conventional probability distribution in the sense that the former allows inexact reasoning at whatever level of abstraction [15] and on the other hand reduces to the latter if basic probabilities are assigned to singleton propositions only. It is in this context that *D–S* theory is claimed to generalise Bayesian inference [25]. Indeed such generalisation differentiates between ignorance (or lack of knowledge,

* Corresponding author.
E-mail addresses: jian-bo.yang@mbs.ac.uk (J.-B. Yang), ling.xu@mbs.ac.uk (D.-L. Xu).

0004-3702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artint.2013.09.003

as defined in Section 2.1) and equal likelihood, and does not assume evidence partially in favour of a proposition to be construed as evidence against the same proposition through the commitment of the remaining belief to its negation [15,20]. These attractive features have motivated the use of *D–S* theory in many areas like knowledge-based system [22,8], pattern recognition [6], information fusion [21,30,14], *Multiple Criteria Decision Analysis* (*MCDA*) and risk analysis [38,40,2,41,37,33,42].

The kernel of *D–S* theory is Dempster's rule [4,5,25], which is rooted in probability theory and constitutes a conjunctive probabilistic inference process. Indeed it generalises Bayes' rule and was promoted as the sole *Evidence Combination Rule* (*ECR*) to combine evidence in the *D–S* framework originally [25]. As it stands however, Dempster's rule has no definition and cannot be applied in a special case when two pieces of evidence are in complete conflict, i.e. with each supporting different propositions. This lack of definition has led to a counter-intuitive problem when the rule is used to combine evidence in high (or near complete) conflict [46,23,16,32]. Another concern about Dempster's rule is that it assumes that each piece of evidence is fully reliable and can veto any proposition, as discussed in Section 2.2 in detail. This means that if a piece of evidence does not support a proposition at all, that proposition will be ruled out completely. In other words, Dempster's rule accumulates consensus support only and rejects a proposition completely if it is opposed by any evidence, no matter what support it may get from any other evidence. While this may be acceptable in special cases, general situations are that multiple pieces of evidence are of a compensatory nature and each play limited roles or have various degrees of reliabilities in support for and opposition against propositions [41,28].

Since the counter-intuitive problem of Dempster's rule was identified over three decades ago, a plethora of alternative *ECR*s have been developed and reviewed in the literature [16,17,14], each with its own merits as well as drawbacks. In this paper, it is not intended to provide another full review of these existing rules. Instead, typical *ECR*s will be briefly compared to emphasise the motivation of this research. Existing alternative *ECR*s are aimed to replace Dempster's rule for addressing the counter-intuitive problem and can be differentiated on the basis of how they deal with conflict among evidence [14]. Three typical views can be found in the literature: (1) allocating conflicting beliefs to the frame of discernment as global ignorance [36], (2) allocating conflicting beliefs to a subset of relevant focal propositions as local ignorance [9] or redistributing it among focal propositions locally [27], and (3) modifying initial belief function to better represent original information without modifying Dempster's rule [16].

One common observation of the alternative *ECR*s is that they are non-probabilistic in the sense that they change the specificity of evidence in basic probability assignment and/or do not constitute a Bayesian inference process when used to combine probability information. It is therefore difficult to interpret the results generated by using the alternative *ECR*s. A critical question then arises as to whether it is meaningful to replace Dempster's rule in situations where it is applicable, or whether it is sufficient to identify rules to combine pieces of highly or completely conflicting evidence. A more general question is how to combine pieces of evidence with various weights and reliabilities that have different meanings. The importance of a piece of evidence reflects a decision maker's preferences over the evidence, which is subjective, depending on who makes the judgement when using the evidence. On the other hand, reliability is used to measure the quality of a piece of evidence objectively, which is the inherent property of the information source where the evidence is generated, and is independent of who may use the evidence [28].

This paper is aimed to address the above questions and establish a unique rule, referred to as *Evidential Reasoning* rule, or *ER* rule for short, to combine multiple pieces of independent evidence for generating their joint support for a proposition. A piece of evidence is said to be independent if the information it carries does not depend on whether other evidence is known or not. This research is also motivated to investigate the rationale and foundation of the *ER* approach, which was developed to support *MCDA* of a quantitative and qualitative nature under uncertainty by applying *D–S* theory [38,40,41,43], to generalise it for evidence combination in general.

In the *ER* approach [41], a basic probability mass is generated by multiplying the degree of belief by the weight of evidence. The *Basic Probability Assignment* (*bpa*) scheme of the *ER* approach ensures that the residual support left uncommitted due to the weight of evidence is made assignable to any singleton propositions and the frame of discernment, depending upon what propositions other evidence supports. In *D–S* theory, the residual support is assigned to a specific proposition: the frame of discernment [25]. This specific assignment does not differentiate between ignorance and the residual support, whilst the former is an intrinsic property of the evidence and the latter reflects its extrinsic feature related to its relative importance compared with other evidence. This indiscrimination changes the specificity of evidence, leading to a dilemma that even if all pieces of evidence point precisely and unambiguously to a proposition their combined support for the proposition generated using the Dempster's rule will still be imprecise or incomplete.

In this paper, the new *ER* rule with evidence weight considered (or *ER* rule with weight in short) is first established by generalising the above *bpa* scheme of the *ER* approach. The first step of the *ER* rule is to construct a new *WBD* as the counterpart of *BD* for a piece of evidence to cater for its extrinsic feature of relative importance. The second step is to implement the orthogonal sum operation to combine the *WBD*s of multiple pieces of independent evidence. *WBD* is then extended to take into account both weight and reliability by constructing a new *WBD* with *Reliability* (*WBDR*). The new *ER* rule with both weight and reliability considered results from implementing the orthogonal sum operation on *WBDR*s. No specificity of any evidence is changed in the process of constructing *WBD* or *WBDR*. The *ER* rule thus constitutes a generic conjunctive probabilistic reasoning process to combine pieces of independent evidence with various weights and reliabilities.

Since it is based on the orthogonal sum operation, the *ER* rule is inherently associative and commutative, meaning that it can be used to combine multiple pieces of evidence in any order without changing the final results. In this paper, it is proven that the *ER* rule satisfies four synthesis axioms that ought to be followed by a rational probabilistic reasoning process [41,17]. It is shown that the *ER* algorithm [41] is a special case of the *ER* rule when the reliability of evidence is equal to its weight that is normalised to be relative to each other among all evidence in question. It is proven that Dempster's rule is also a special case of the *ER* rule when each piece of evidence is fully reliable. In particular, the *ER* rule completes and enhances Dempster's rule by identifying, through a novel reliability perturbation analysis proposed in this paper, how to combine multiple pieces of independent evidence that are each fully reliable but highly or completely conflicting with each other. The *ER* rule is compared with other *ECR*s that can be used to combine pieces of evidence with different weights and reliabilities, in particular the *PCR5* rule [28]. Comprehensive numerical studies are conducted to validate and illustrate the potential applications of the *ER* rule in comparison with other *ECR*s.

The rest of the paper is organised as follows. In Section 2, the concepts of *D–S* theory and Dempster's rule are briefly introduced, and several other typical *ECR*s are discussed. In Section 3, following an investigation into evidence weight and a brief introduction to both the *PCR5* rule and the *ER* approach, the rationale and process to construct *WBD* are investigated, leading to the establishment of the new *ER* rule with weight. A numerical study is conducted to compare the *ER* rule with other *ECR*s. Section 4 is dedicated to the establishment of the new *ER* rule with both weight and reliability considered, and a comprehensive numerical study is conducted to validate and compare the *ER* rule with other *ECR*s. The properties of the *ER* rule are explored in Section 5, where it is proven that both the *ER* approach and Dempster's rule are special cases of the *ER* rule. The reliability perturbation analysis of the *ER* rule is reported, resulting in the completion and enhancement of Dempster's rule for combination of evidence in high or complete conflict. The simulation study is conducted in this section to show how the *ER* rule can be used for conflict resolution with Zadeh's example revisited and analysed in detail. The paper is concluded in Section 6.

2. Brief introduction to *D–S* theory and typical *ECR*s

The establishment of the new *ER* rule is based on *D–S* theory, in particular the concept of *BD* and the orthogonal sum operation. It is thus necessary to provide a brief introduction to the *prior* knowledge as a basis for later discussions.

2.1. Basic concepts of D–S theory

D–S theory was originally investigated in the 1960's [4,5], formalised in the 1970's [25] and has been researched ever since [3,36,26,29,38,23,41,16,33,34,14,8]. In this subsection, the basic concepts of *D–S* theory are briefly discussed.

Suppose $\Theta = \{\theta_1, \ldots, \theta_N\}$ is a set of mutually exclusive and collectively exhaustive propositions, with $\theta_i \cap \theta_j = \emptyset$ for any $i, j \in \{1, \ldots, N\}$ and $i \neq j$ where $\emptyset$ is an empty set. Θ is then referred to as a frame of discernment. A *Basic Probability Assignment* (*bpa*) is a function $m : 2^\Theta \to [0, 1]$, satisfying [25]

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{\theta \subseteq \Theta} m(\theta) = 1 \tag{1}$$

with 2^Θ or $P(\Theta)$ representing the power set of Θ, consisting of the 2^N subsets of Θ, or

$$P(\Theta) = 2^\Theta = \big\{\emptyset, \theta_1, \ldots, \theta_N, \{\theta_1, \theta_2\}, \ldots, \{\theta_1, \theta_N\}, \ldots, \{\theta_1, \ldots, \theta_{N-1}\}, \Theta\big\} \tag{2}$$

$m(\theta)$ is a basic probability mass that is assigned exactly to a proposition θ and to no smaller subset. Basic probability masses assigned to all the subsets of Θ are summed to unity and there is no belief left to the empty set. In this paper, a basic probability mass assigned exactly to Θ is referred to as the degree of *global ignorance*, denoted by $m(\Theta)$; a basic probability mass assigned exactly to a smaller subset of Θ except for any singleton proposition or Θ is referred to as the degree of *local ignorance*. If there is no local or global ignorance, a *bpa* function will reduce to a classical probability function.

Associated with each *bpa* are a belief measure, denoted by $Bel(\theta)$, and a plausibility measure, denoted by $Pl(\theta)$, which are defined by the following equations [25]:

$$Bel(\theta) = \sum_{B \subseteq \theta} m(B) \quad \text{and} \quad Pl(\theta) = \sum_{B \cap \theta \neq \emptyset} m(B) \tag{3}$$

$Bel(\theta)$ represents all basic probability masses assigned exactly to θ and its smaller subsets, and $Pl(\theta)$ represents all possible basic probability masses that could be assigned to θ and its smaller subsets. As such, $Bel(\theta)$ and $Pl(\theta)$ can be interpreted as the lower and upper bounds of probability to which θ is supported [36,10,31]. The two measures are connected by the following equation [25]

$$Pl(\theta) = 1 - Bel(\bar{\theta}) \tag{4}$$

where $\bar{\theta}$ denotes the complement or negation of θ. The difference between the belief and plausibility measures of θ describes the degree of ignorance for θ.

2.2. Dempster's rule

The kernel of *D–S* theory is Dempster's rule that was originally adopted as the sole *ECR* to combine independent evidence that is fully reliable. By fully reliable evidence, it is meant that a proposition will not be supported at all if it is ruled out by the evidence. The rule adopts the orthogonal sum operation to combine evidence, which is rooted in calculating the joint probability of independent events [19]. With two pieces of independent and fully reliable evidence represented by two *bpas* m_1 and m_2 respectively, for any proposition $\theta \subseteq \Theta$, Dempster's rule is given as follows [25]

$$m(\theta) = [m_1 \oplus m_2](\theta) = \begin{cases} 0 & \theta = \emptyset \\ \frac{\sum_{B \cap C = \theta} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)} & \theta \neq \emptyset \end{cases} \tag{5}$$

where $\oplus$ is the orthogonal sum operator.

Dempster's rule forms a conjunctive probabilistic reasoning process, takes Bayes' rule as a special case, and is both associative and commutative [25,19], which are among the most important and useful properties for both human and machine reasoning. It is clear from the multiplication operation in the above equation that Dempster's rule provides a process for combining two pieces of non-compensatory evidence, in the sense that if either of them completely opposes a proposition, the proposition will not be supported at all, no matter how strongly it may be supported by the other piece of evidence. For example, if $m_1(B) = 0$ for any $B \cap \theta = \theta$, meaning that θ is not supported by m_1 at all, the numerator of formula (5) will be zero. In other words, in Dempster's rule, a proposition will be supported only if both pieces of evidence each support it to some degrees. Note that Dempster's rule is not defined where two pieces of evidence are in complete conflict, or $\sum_{B \cap C = \emptyset} m_1(B) m_2(C) = 1$, as in this case the denominator of formula (5) is zero, as is the numerator. It is shown that Dempster's rule may lead to a counter-intuitive result when used to combine highly conflicting evidence [45,46,23,32]. As mentioned in the literature, because the two pieces of evidence in Zadeh's example [45,46] are both Bayesian and Bayes's Rule is a special case of Dempster's rule [25], such a counter-intuitive problem is also associated with classical probabilistic reasoning [16]. It is therefore fundamental to deal with the non-definition and counter-intuitive problem in the context of probabilistic reasoning.

2.3. Other typical ECRs for comparison

To overcome the non-definition and counter-intuitive problem associated with Dempster's rule, a plethora of different *ECRs* have been developed on the basis of various views. Three different views were outlined in the introduction section. The rest of this section is dedicated to discussing these different views in some detail.

Yager's rule [36] can be described by the following equations:

$$m(\theta) = \begin{cases} \sum_{B \cap C = \theta} m_1(B) m_2(C) & \theta \subset \Theta, \theta \neq \emptyset \\ 0 & \theta = \emptyset \\ m_1(\theta) m_2(\theta) + \sum_{B \cap C = \emptyset} m_1(B) m_2(C) & \theta = \Theta \end{cases} \tag{6}$$

Different from Dempster's rule, Yager's rule is proposed to allocate conflicting beliefs ($\sum_{B \cap C = \emptyset} m_1(B) m_2(C)$ of Eq. (6)) to the frame of discernment, thus avoiding the non-definition and counter-intuitive problem of Dempster's rule. However, this solution comes with high prices. On one hand, it treats two pieces of completely conflicting evidence in the same way as if there was no evidence at all. On the other hand, because of its way to allocate conflicting beliefs, it does not constitute a conjunctive reasoning process or the orthogonal sum operation as Dempster's rule does, so it is not a probabilistic reasoning process. In other words, the use of Yager's rule to combine two pieces of evidence in conflict but with no ignorance, as represented by a standard probability distribution, will result in a combined distribution that is not a probability distribution in the sense that it will have global ignorance. A question then arises as to whether such treatment of conflict is rational or not, as it introduces ignorance which "invents" uncertainty unnecessarily and could be misleading.

Dubois and Prade's rule [9] can be described as follows

$$m(\theta) = \begin{cases} 0 & \theta = \emptyset \\ \sum_{B \cap C = \theta} m_1(B) m_2(C) + \sum_{\substack{B \cup C = \theta \\ B \cap C = \emptyset}} m_1(B) m_2(C) & \theta \neq \emptyset \end{cases} \tag{7}$$

Dubois and Prade's rule can also avoid the non-definition and counter-intuitive problem associated with Dempster's rule. Compared with Yager's rule, it allocates conflicting beliefs locally in the sense that conflict should be resolved among focal propositions in which the conflict occurs. This is perhaps a more rational view than treating any conflict as global ignorance as in Yager's rule. However, Dubois and Prade's rule does not constitute a conjunctive or probabilistic reasoning process either, in the sense that there is no probabilistic rule to legitimise the conjunctive belief $m_1(B) m_2(C)$ for θ from the union $B \cup C = \theta$ with $B \cap C = \emptyset$. If Dubois and Prade's rule is used only to replace Dempster's rule for combining highly or completely conflicting evidence, there are issues about how to combine evidence with or without high conflict consistently and how to distribute beliefs among focal propositions for conflict resolution, which will be investigated in next sections.

The *PCR5* rule [27] evolved from Dubois and Prade's rule and can be described as follows

$$m(\theta)=\begin{cases}0 & \theta=\emptyset\\ \sum_{B\cap C=\theta}m_1(B)m_2(C)+\sum_{\theta\cap B=\emptyset}[\frac{m_1(\theta)^2m_2(B)}{m_1(\theta)+m_2(B)}+\frac{m_2(\theta)^2m_1(B)}{m_2(\theta)+m_1(B)}] & \theta\neq\emptyset\end{cases}\tag{8}$$

If the denominator is zero, the fraction will not be taken into account. From Eq. (8), it is clear that apart from inheriting the merits of Dubois and Prade's rule the *PCR5* rule also proposes a way to redistribute conflicting beliefs locally. Nonetheless, unlike Dempster's rule, the *PCR5* rule is not a conjunctive or probabilistic reasoning process, nor is it Bayesian reasoning process when used to combine evidence with no ignorance. Indeed it is termed as a plausible and paradoxical reasoning process by its authors [13,27]. If the *PCR5* rule is used for conflict resolution only, there is an issue of how to interpret and justify the way of proportionally redistributing conflicting beliefs among the focal propositions concerned. This issue will be investigated in next sections.

The third view [16] is not to create alternative rules for replacing Dempster's rule in situations where it is applicable. Rather, it is only necessary to fix the problem of Dempster's rule when used to combine highly or completely conflicting evidence. However, the two approaches proposed by Haenni [16] are either to assume that highly conflicting evidence could be interpreted and represented in different ways or that experts are not fully reliable. While these approaches are aimed to reduce conflict at the modelling stage, they lead to changing the specificity of evidence but still do not resolve the problem of how to consistently combine pieces of fully reliable evidence that are genuinely in high or complete conflict, or how to ensure that such a combination process still constitutes a probabilistic reasoning process and keeps intact the most important property of Dempster's rule that it generalises Bayes' rule. These different views will be further investigated in next sections.

3. New *ER* rule to combine evidence with weight

To deal with the non-definition and counter-intuitive problem of Dempster's rule and answer the questions discussed above, it is helpful to review the origin of *D–S* theory as described in Shafer's book [25]. As explained at the start of the first chapter of the book, *D–S* theory "is a theory of evidence because it deals with weights of evidence and with numerical degree of support based on evidence". Here the importance of and the difference between the concepts of evidence weight and evidence support cannot be made clearer. As such, it is fundamental to start with investigating the key issue of how to deal with evidence weight in an evidence combination process. A few relevant *ECR*s are then examined in this section.

3.1. Evidence weight

What is the weight of evidence and how to quantify it? In Shafer's book, this question was first analysed in an intuitive and straightforward manner, though followed by a chapter (Chapter 5) dedicated to this topic in a rather abstract or theoretical way. Nevertheless, it is the intuitive explanation of evidence weight as described in Section 1 on Simple Support Function in Chapter 4 of Shafer's book [25] that provides an important hint to answer this question. In that section, it is stated that in a situation "where the evidence points precisely and unambiguously to a single non-empty subset A of Θ", "we can say that the effect of the evidence is limited to providing a certain degree of support for A", denoted by $S(A)$ with $0\leqslant S(A)\leqslant 1$. For such a statement, Shafer used a term "simple support function" to name $S(A)$ by saying at the end of this section that "indeed, if $S(A)$ is a simple support function focused on A, then S is the belief function with the basic probability numbers $m(A)=S(A)$, $m(\Theta)=1-S(A)$ and $m(B)=0$ for all other $B\subset\Theta$".

Based on the above arguments, one cannot help but comprehend that $S(A)$ has already taken into account the weight of evidence, just as mentioned at the beginning of Chapter 5 in Shafer's book where it is said that weights of evidence were translated into simple support functions as discussed in Chapter 4. Following these arguments, let w stand for the weight of a piece of evidence and $p(\theta)$ for the degree of belief to which the evidence points to a single non-empty subset θ of Θ, or the probability to which the proposition θ can be proven when the evidence is gathered, we can have the following relationship among support function $S(\theta)$, weight w and the degree of belief $p(\theta)$

$$m(\theta)=S(\theta)=wp(\theta)\quad\text{with } 0\leqslant w\leqslant 1,\ 0\leqslant p(\theta)\leqslant 1 \text{ and } \sum_{\theta\subseteq\Theta}p(\theta)=1\tag{9}$$

What Eq. (9) means is that the degree of support for a proposition is proportional to both the weight of the evidence and the belief degree to which the evidence points to the proposition. In Shafer's simple support function, $p(\theta)$ is assumed to be one. From Eq. (9), we then get $m(\theta)=w$. As investigated in the next two sections, the above ideas about evidence weight have been embedded in some existing methods. The difference is how to handle the residual support left by the evidence due to its weight, as measured by $1-w$. In Shafer's book, the residual support is allocated to Θ, or $m(\Theta)=1-S(\theta)=1-w$, as discussed above. In literature, this idea is generalised and formalised as Shafer's discounting method, defined for convenience of discussion as follows.

Definition 1 *(Shafer's discounting).* Suppose $p(\theta)$ is the degree of belief to which a piece of evidence points to a proposition θ. Let α be a factor that is used to discount $p(\theta)$, where α $(0\leqslant\alpha\leqslant 1)$ can be interpreted as the reliability or importance of the evidence, depending upon circumstances in which α is used. Then, Shafer's discounting method is defined to generate *bpa* for the evidence as follows

$$m(\theta)=\begin{cases}\alpha p(\theta) & \theta\subset\Theta,\theta\neq\emptyset\\ \alpha p(\theta)+(1-\alpha) & \theta=\Theta\\ 0 & \theta=\emptyset\end{cases}\qquad \text{with } 0\leqslant\alpha\leqslant 1 \tag{10}$$

Shafer's discounting method changes the specificity of the original evidence in the sense that as long as $1-\alpha>0$, global ignorance is introduced to a belief distribution even when the evidence points to a proposition θ precisely and unambiguously, or $p(\theta)=1$. In other words, if $\alpha<1$ there will always be $m(\Theta)\geqslant 1-\alpha>0$ even if $p(\theta)=1$. In what follows, two other schemata for allocating the residual support $1-\alpha$ are examined and the rationale of using Eq. (9) to assign basic probability mass or the degree of support for a proposition is discussed in the context of information fusion and multiple criteria decision analysis, both of which share some common features and requirements.

3.2. Introduction to the PCR5 rule with evidence weight

PCR5 was extended to deal with evidence weight in the context of importance discounting [28]. In this approach, the weight of a source is characterised by an importance factor, or w with $0\leqslant w\leqslant 1$. In the extended *PCR5*, the degree of support for a proposition is given in the same way as in Eq. (9), but the residual support is assigned to the empty set as shown in the follow definition.

Definition 2 *(PCR5 importance discounting).* Suppose w is the weight of a piece of evidence. The *PCR5* importance discounting method is defined to generate *bpa* for the evidence as follows

$$m(\theta)=\begin{cases}wp(\theta) & \theta\subseteq\Theta,\theta\neq\emptyset\\ wp(\theta)+(1-w) & \theta=\emptyset\end{cases}\qquad \text{with } 0\leqslant w\leqslant 1 \tag{11}$$

It is important to note that in Eq. (11) *bpa* is assigned to the empty set. This means that the distribution generated using Eq. (11) is not a probability distribution even when there is no ignorance in the original evidence, nor is it a belief distribution as defined in *D–S* theory because similar to probability theory no belief can be assigned to the empty set in *D–S* theory. Indeed, a new theory coined as *DSmT* [27] is proposed to deal with such *bpa*s. In *DSmT*, the *PCR5* rule is modified to handle *bpa*s as generated by Eq. (11), referred to as $PCR5_{\emptyset}$ which is the same as Eq. (8) except that $m(\emptyset)$ is not set to zero and the empty set is treated in the same way as any other subsets of Θ in the process of evidence combination. This is the only difference between $PCR5_{\emptyset}$ and *PCR5*. While this paper is not intended to investigate *DSmT* in detail, it is worth noting that $PCR5_{\emptyset}$ does not constitute a probabilistic reasoning process for evidence combination as discussed in *D–S* theory or in classical Bayesian probability theory. As such, there is a question as to how to interpret the results generated by applying $PCR5_{\emptyset}$. Another concern is that because the $PCR5_{\emptyset}$ rule does not have the property of commutativity, the final results generated by the $PCR5_{\emptyset}$ rule depend on the order in which evidence is combined. In *DSmT*, much attention is paid to simultaneous combination of evidence.

3.3. Introduction to the ER approach and its weight handling method

In the *ER* approach [38–41,44], the assessment of an alternative on each criterion is regarded as a piece of evidence; the weight of evidence is equal to the weight of a criterion. Suppose there are a number of alternatives and each is assessed on L criteria e_i $(i=1,\dots,L)$ using a common set of N assessment grades (propositions) θ_n $(n=1,\dots,N)$, constituting a frame of discernment denoted by $\Theta=\{\theta_n,\ n=1,\dots,N\}$. If an alternative is assessed to a grade θ_n on a criterion e_i with a belief degree $p_{n,i}$, this assessment can be profiled as a BD defined by:

$$e_i=\big\{(\theta_n,p_{n,i}),\ n=1,\dots,N;\ (\Theta,p_{\Theta,i})\big\} \tag{12}$$

with $0\leqslant p_{n,i}\leqslant 1$ $(n=1,\dots,N)$, $\sum_{n=1}^{N}p_{n,i}\leqslant 1$ and $p_{\Theta,i}=1-\sum_{n=1}^{N}p_{n,i}$ being the degree of global ignorance.

Suppose w_i is the weight of the ith criterion, normalised to represent the relative importance of the criterion by

$$0\leqslant w_i\leqslant 1\quad \text{for } i=1,\dots,L \text{ and } \sum_{i=1}^{L}w_i=1 \tag{13}$$

The *ER* algorithm can now be briefly summarised and implemented recursively. First, the basic probability masses for e_1 are generated as weighted belief degrees as follows

$$m_{n,1}=w_1p_{n,1}\quad \text{for } n=1,\dots,N,\qquad m_{\Theta,1}=w_1p_{\Theta,1},\quad \text{and}\quad m_{P(\Theta),1}=1-w_1 \tag{14}$$

Note that $m_{P(\Theta),1}$ is the remaining support left uncommitted by e_1, coined as the *residual support* of e_1 that cannot be assigned by e_1 alone due to its weight. It is assignable to any individual and/or subsets of grades, depending upon the nature and weights of other pieces of evidence. Similarly, basic probability masses for another assessment e_2 are generated by

$$m_{n,2} = w_2 p_{n,2} \quad \text{for } n = 1, \dots, N, \qquad m_{\Theta,2} = w_2 p_{\Theta,2}, \quad \text{and} \quad m_{P(\Theta),2} = 1 - w_2 \tag{15}$$

The basic probability masses of e_1 and e_2 are combined into the joint probability masses for θ_n, Θ and $P(\Theta)$ using the following equations:

$$\{\theta_n\}: \quad m_{n,e(2)} = k\big(m_{n,1}m_{n,2} + m_{n,1}(m_{\Theta,2} + m_{P(\Theta),2}) + (m_{\Theta,1} + m_{P(\Theta),1})m_{n,2}\big) \quad n = 1, \dots, N \tag{16a}$$

$$\{\Theta\}: \quad m_{\Theta,e(2)} = k(m_{\Theta,1}m_{\Theta,2} + m_{\Theta,1}m_{P(\Theta),2} + m_{P(\Theta),1}m_{\Theta,2}) \tag{16b}$$

$$\big\{P(\Theta)\big\}: \quad m_{P(\Theta),e(2)} = k(m_{P(\Theta),1}m_{P(\Theta),2}) \tag{16c}$$

$$k = \left(1 - \sum_{n=1}^{N} \sum_{t=1,\, t\neq n}^{N} m_{n,1}m_{t,2}\right)^{-1} \tag{16d}$$

If there are more than two assessments to be combined, Eqs. (16a) to (16d) can be repeated to combine the 3rd assessment with the previously-combined assessment $m_{n,e(2)}$ ($n = 1, \dots, N$), $m_{\Theta,e(2)}$ and $m_{P(\Theta),e(2)}$, and so on until all assessments are combined recursively. The combined belief degrees, denoted by $p_{n,e(2)}$ and $p_{\Theta,e(2)}$ for $L = 2$ without loss of generality, are generated by reassigning $m_{P(\Theta),e(2)}$ back to all focal elements of Θ as follows:

$$\{\theta_n\}: \quad p_{n,,e(2)} = \frac{m_{n,e(2)}}{1 - m_{P(\Theta),e(2)}}, \quad n = 1, \dots, N \tag{17a}$$

$$\{\Theta\}: \quad p_{\Theta,e(2)} = \frac{m_{\Theta,e(2)}}{1 - m_{P(\Theta),e(2)}} \tag{17b}$$

The combined assessment is then profiled by the following combined *BD*:

$$e(2) = \big\{(\theta_n, p_{n,e(2)}),\ \text{for } n = 1, \dots, N;\ (\Theta, p_{\Theta,e(2)})\big\} \tag{18}$$

The combined *BD* provides a panoramic view about the overall performance of an alternative with the degrees of its strengths and weaknesses (represented by the assessment grades θ_n) explicitly measured by the belief degrees. We now show that the above-outlined original *ER* algorithm can be rewritten in the same way as given by the following lemma.

Lemma 1. *The combined degrees of belief $p_{n,e(2)}$ and $p_{\Theta,e(2)}$ generated using the original ER algorithm of Eqs. (12) to (18) can be equivalently rewritten as follows*:

$$p_{n,e(2)} = k_1 \hat{m}_{n,e(2)} \quad \textit{for } n = 1, \dots, N, \quad \textit{and} \quad p_{\Theta,e(2)} = k_1 \hat{m}_{\Theta,e(2)} \tag{19a}$$

$$k_1 = \left(\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}\right)^{-1} \tag{19b}$$

$$\hat{m}_{n,e(2)} = \big[(1 - w_2)m_{n,1} + (1 - w_1)m_{n,2}\big] + [m_{n,1}m_{n,2} + m_{n,1}m_{\Theta,2} + m_{\Theta,1}m_{n,2}] \tag{19c}$$

$$\hat{m}_{\Theta,e(2)} = \big[(1 - w_2)m_{\Theta,1} + (1 - w_1)m_{\Theta,2}\big] + [m_{\Theta,1}m_{\Theta,2}] \tag{19d}$$

Proof. See Appendix A.1. □

In Eq. (19c), there are two square bracket terms. We name the first term as the ***bounded sum of individual support*** from each of the two pieces of evidence e_1 and e_2 for the proposition that an alternative is assessed to θ_n. In other words, it is the sum of the degree of individual support $m_{n,1}$ from e_1 bounded by $(1 - w_2)$ and the degree of the individual support $m_{n,2}$ from e_2 bounded by $(1 - w_1)$. Note that w_1 is the weight of e_1 and $(1 - w_1)$ sets a bound on the role that e_2 can play in its individual support for a proposition. The first term implies that the higher the importance of a piece of evidence, the less room it leaves for the other evidence to play. We name the second square bracket term as the ***orthogonal sum of collective support*** from both e_1 and e_2, with $m_{n,1}m_{n,2}$ measuring the degree of direct or homogeneous support and $(m_{n,1}m_{\Theta,2} + m_{\Theta,1}m_{n,2})$ measuring the degree of intersected or heterogeneous support for the proposition that an alternative is assessed to θ_n due to the fact that $\theta_n \cap \Theta = \theta_n$ for any $n = 1, \dots, N$ and $\theta \cap P(\Theta) = \theta$ for any $\theta \subseteq \Theta$.

The above discussions show that the combined degree of belief $p_{n,e(2)}$ given by Eq. (19a) is equal to the bounded sum of individual support plus the orthogonal sum of collective support from e_1 and e_2 for a proposition that an alternative is assessed to θ_n. This is an interesting result that we will discuss in more detail in next sections.

3.4. New weighted belief distribution (WBD)

In Eq. (12), only global ignorance was taken into account. In general, a piece of evidence e_i can be profiled by a *BD*, defined as follows

$$e_i = \left\{(\theta, p_{\theta,i}),\ \forall\theta \subseteq \Theta,\ \sum_{\theta\subseteq\Theta} p_{\theta,i} = 1\right\} \tag{20}$$

where $(\theta, p_{\theta,i})$ is an element of evidence e_i, representing that the evidence points to proposition θ to the degree of $p_{\theta,i}$. $(\theta, p_{\theta,i})$ will be referred to as a focal element of e_i if $p_{\theta,i} > 0$. In Eq. (20), both global ignorance and local ignorance are taken into account. Suppose the weight of evidence e_i is denoted by w_i for $i = 1, \ldots, L$, which are not necessarily normalised to unity as required in Eq. (13).

We are now in a position to generalise the *bpa* calculation method proposed in the *ER* approach, as shown by Eq. (14), for assigning basic probability masses generally. First of all, it is worth noting that the method shown in Eq. (14) recognises the difference between residual support and the degree of global ignorance. In other words, the residual support of e_i is measured by $m_{P(\Theta),i} = 1 - w_i$, which is not assigned to Θ as global ignorance and hence is not part of $m_{\Theta,i}$ in e_i in both *bpa* calculation and the process of combining evidence. The importance of this distinction cannot be overemphasised because the global ignorance in a piece of evidence is its intrinsic property and has no relevance to other evidence, whilst the residual support is the extrinsic feature of the evidence and is incurred due to the relative importance of the evidence compared with other evidence. Without making this distinction, global ignorance would be introduced into a piece of evidence even if the evidence originally has no ignorance at all, which would inevitably lead to the change of the specificity or the distortion of the evidence. Combining distorted evidence leads to irrational results in the sense that even if all pieces of evidence are given as precise probability distributions with no ignorance at all their combined results will still have a degree of global ignorance.

If such distinction is ignored, for example, as in the case of using Shafer's discounting method to take into account weights of evidence, we would have to face the question of how to interpret the results generated. To illustrate the above arguments, suppose the following two pieces of evidence are given equal weight $w_1 = w_2 = 1 - \gamma$ with $0 \leqslant \gamma \leqslant 1$

$$e_1 = \{(A, 1-\delta), (B, \delta)\} \quad \text{and} \quad e_2 = \{(B, \delta), (C, 1-\delta)\} \tag{21}$$

where A, B and C are mutually exclusive and collectively exhaustive propositions, and δ is a small positive real number with $0 \leqslant \delta \leqslant 1$.

If δ is set to be a small number, for example $\delta = 0.01$, and Shafer's method is used to discount the two pieces of evidence by their weights w_1 and w_2, we would have the following discounted *BD*s

$$m_1 = \big\{\big(A, 0.99(1-\gamma)\big), \big(B, 0.01(1-\gamma)\big), \big(\{A, B, C\}, \gamma\big)\big\} \tag{22}$$

$$m_2 = \big\{\big(B, 0.01(1-\gamma)\big), \big(C, 0.99(1-\gamma)\big), \big(\{A, B, C\}, \gamma\big)\big\} \tag{23}$$

where the element $\{A, B, C\}$ is the frame of discernment representing global ignorance.

If γ is taken as a small number, for example $\gamma = 0.05$, applying Dempster's rule to aggregate the above two discounted *BD*s (22) and (23) will lead to the following result:

$$m = m_1 \oplus m_2 = \big\{(A, 0.4819), (B, 0.0107), (C, 0.4819), \big(\{A, B, C\}, 0.0256\big)\big\} \tag{24}$$

This result makes sense but is not precise, because the aggregated result of the two pieces of evidence e_1 and e_2 with no ignorance at all still includes a belief degree of 0.0256 allocated to the frame of discernment showing a degree of global ignorance, even though the degree 0.0256 is rather small.

It is easy to show that for any $\gamma > 0$ the result generated by combining (22) and (23) using Dempster's rule will always lead to similar imprecise results in the sense that the combined degree of belief in $\{A, B, C\}$ will always be positive. An extreme case in the above example is to set $\delta = 1$ and $\gamma > 0$. In this case, B is surely supported by each of the two pieces of evidence with no ignorance or ambiguity; however, their combination leads to a positive degree of belief in $\{A, B, C\}$ for any $\gamma > 0$, which is obviously irrational. This will still be the case even if there are many more pieces of evidence surely supporting B.

The above discussions show that it is irrational to assign the residual support specifically to the frame of discernment. Following similar arguments, it can be asserted that it is irrational to assign the residual support specifically to only one of other propositions. If Eq. (1) is followed, as is the case in *D–S* theory, the residual support must not be assigned to the empty set either. So, it seems that the only appropriate way to allocate the residual support is to make it assignable to any propositions with no *prior* specification, which implies that the residual support should be assigned to the power set of the frame of discernment instead of specifically to the frame of discernment or any its single subset. The above discussion leads to the following definition.

Definition 3 *(Weighted belief degree).* Suppose w_i $(0 \leqslant w_i \leqslant 1)$ is the weight of evidence e_i defined in Eq. (20) with $w_i = 0$ and 1 standing for "not important at all" and "the most important" respectively. The weighted belief degree for e_i are then defined as follows

$$m_{\theta,i} = m_i(\theta) = \begin{cases} 0 & \theta = \emptyset \\ w_i p_{\theta,i} & \theta \subseteq \Theta, \theta \neq \emptyset \\ 1 - w_i & \theta = P(\Theta) \end{cases} \tag{25}$$

The term $m_{\theta,i}$ may also be referred to as basic probability mass or the degree of support for θ from e_i. Definition 3 generalises the *bpa* method of the *ER* approach as shown in Eq. (14) to handle both local and global ignorance. It is worth noting that multiplying $p_{\theta,i}$ by the same weight w_i for all $\theta \subseteq \Theta$ does not change the relativity property of belief degree, or $m_{A,i}/m_{B,i} = p_{A,i}/p_{B,i}$ for any $A, B \subseteq \Theta$. As such, the specificity of evidence e_i is kept intact. In Definition 3 the residual support $(1 - w_i)$ of e_i is nominally attached to $P(\Theta)$ without being pre-assigned to any specific proposition of Θ. The residual support can thus be conjunctively redistributed to all propositions because $\theta \cap P(\Theta) = \theta$ for any $\theta \subseteq \Theta$. This can ensure genuine conjunctive reasoning in an evidence combination process.

To facilitate conjunctive reasoning based on Definition 3, evidence profiled by Eq. (20) needs to be represented by a so-called *Weighted Belief Distribution* (*WBD*) defined as follows. First, note from the above discussions that the residual support $(1 - w_i)$ of e_i is the amount of remaining support left uncommitted by e_i due to its weight. This means that e_i plays only a limited role equal to its weight w_i in support for a proposition and the quantity $(1 - w_i)$ sets a bound on the role that another piece of evidence can play with which e_i is combined. The *WBD* of e_i, denoted by m_i, is then constructed from its *BD* by replacing each belief degree $p_{\theta,i}$ with its weighted belief degree $m_{\theta,i}$, and by adding to *BD* a special element $(P(\Theta), m_{P(\Theta),i})$ as follows

$$m_i = \big\{(\theta, m_{\theta,i}),\ \forall\theta \subseteq \Theta;\ \big(P(\Theta), m_{P(\Theta),i}\big)\big\} \tag{26}$$

Note that $m_{P(\Theta),i} = 1 - w_i$ is the residual support of evidence e_i incurred due to its weight and there is $\sum_{\theta\subseteq\Theta} m_{\theta,i} + m_{P(\Theta),i} = w_i \sum_{\theta\subseteq\Theta} p_{\theta,i} + (1 - w_i) = 1$ since $\sum_{\theta\subseteq\Theta} p_{\theta,i} = 1$ in Eq. (20). It is interesting to note from Definition 3 that the following one-to-one connection always holds between a *BD* and its corresponding *WBD*:

$$p_{\theta,i} = \frac{m_{\theta,i}}{1 - m_{P(\Theta),i}}, \quad \forall\theta \subseteq \Theta \tag{27}$$

The above analysis shows that a *BD* (Eq. (20)) and its *WBD* (Eq. (26)) are complementary to each other for representing the same piece of evidence and can be deduced from each other through Eqs. (25) and (27). In other words, *BD* and *WBD* are the two sides of the same coin for representing the same piece of evidence. While the *BD* of a piece of evidence is used to describe its intrinsic property of how it points to propositions, its corresponding *WBD* is used to represent its extrinsic support for propositions in comparison with other evidence, as discussed in detail later in this paper. A special case of Eq. (26) is when there is no local ignorance. In this case, *WBD* reduces to the following format:

$$m_i = \big\{(\theta_n, w_i p_{\theta_n,i}),\ n = 1, \ldots, N;\ (\Theta, w_i p_{\Theta,i});\ \big(P(\Theta), (1 - w_i)\big)\big\} \tag{28}$$

Note that in Eq. (28) there is $\sum_{n=1}^{N} w_i p_{\theta_n,i} + w_i p_{\Theta,i} + (1 - w_i) = 1$ as $\sum_{n=1}^{N} p_{\theta_n,i} + p_{\Theta,i} = 1$. The original *ER* algorithm is generated by implementing the orthogonal sum operation on *WBD*s, as given by the following theorem.

Theorem 1 *(Original ER algorithm). The ER algorithm given by Eqs.* (12)–(18) *is the result of implementing the orthogonal sum operation on WBDs defined by Eq.* (28) *with* $\sum_{i=1}^{L} w_i = 1$.

Proof. See Appendix A.2. □

Theorem 1 asserts that to generate the combined belief degrees from two pieces of independent evidence with their relative weights, the orthogonal sum operation should be applied to combine their *WBD*s rather than *BD*s. This is logical as the former takes into account interrelationships among evidence whilst the latter does not.

3.5. New ER rule with evidence weight

In Theorem 1, we assumed that there exists no local ignorance in all evidence. In the following discussions, we drop this assumption by taking into account both global and any local ignorance. We thus generalise the *ER* algorithm to the following *ER* rule.

Theorem 2 *(ER rule with weight). Suppose two pieces of independent evidence* e_1 *and* e_2 *are each profiled by Eq.* (20) *and their WBDs are represented by Eqs.* (25) *and* (26)*. The combined degrees of belief to which both* e_1 *and* e_2 *jointly support proposition* θ*, denoted by* $p_{\theta,e(2)}$*, is given as follows*

$$p_{\theta,e(2)} = \begin{cases} 0 & \theta = \emptyset \\ \frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)}} & \theta \subseteq \Theta, \theta \neq \emptyset \end{cases} \tag{29}$$

$$\hat{m}_{\theta,e(2)} = \left[(1-w_2)m_{\theta,1} + (1-w_1)m_{\theta,2}\right] + \sum_{B\cap C=\theta} m_{B,1}m_{C,2} \quad \forall\theta \subseteq \Theta \tag{30}$$

with w_1 and w_2 as the weights of evidence e_1 and e_2 respectively which are not necessarily normalised, $0 \leqslant p_{\theta,e(2)} \leqslant 1$ $\forall\theta \subseteq \Theta$ and $\sum_{\theta\subseteq\Theta} p_{\theta,e(2)} = 1$.

Proof. See Appendix A.3. □

From the proof of Theorem 2, it can be noted that $\hat{m}_{\theta,e(2)}$ measures the amount of total joint support for proposition θ from both e_1 and e_2, generated by the conjunction or the orthogonal sum of the two *WBDs* for e_1 and e_2. Note that the conjunction of a proposition θ with the power set $P(\Theta)$ is still itself θ. As such, for example, the residual support $(1-w_1)$ of evidence e_1 is conjunctively redistributed to all the focal elements of e_2. This ensures that the specificity of a belief distribution is kept intact in the evidence combination process in the sense that only the focal elements of e_1 and e_2 will remain as focal elements after combination. Consequently, in the combination process no belief is distributed to the empty set that is not a focal element in any legitimate belief distribution.

Theorem 2 reinforces the notion that the combined degree of belief to which two pieces of independent evidence jointly support a proposition consists of two parts: the bounded sum of their individual support (the square bracket term in Eq. (30)) and the orthogonal sum of their collective support (the last term of Eq. (30)). The rationale of Eq. (30) can be explained as follows. If two pieces of evidence each play limited roles bounded by their weights, in addition to their collective support, individual support from any evidence should be counted as part of the combined support in general. If one piece of evidence plays a dominant role, individual support from the other evidence will be counted only to reinforce what are already supported by the dominant evidence but not for any other propositions. To combine more than two pieces of independent evidence, we have the following results.

Corollary 1 *(Recursive combination of WBDs). Suppose L pieces of independent evidence are each profiled by Eq. (20) and their WBDs are represented by Eqs. (25) and (26). Suppose $e(i)$ denotes the combination of the first i pieces of evidence and $m_{\theta,e(i)}$ is the probability mass to which θ is supported jointly by $e(i)$, with $m_{\theta,e(1)} = m_{\theta,1}$ and $m_{P(\Theta),e(1)} = m_{P(\Theta),1}$. The orthogonal sum of the first i WBDs is then given as follows*

$$m_{\theta,e(i)} = [m_1 \oplus \cdots \oplus m_i](\theta) = \begin{cases} 0 & \theta = \emptyset \\ \frac{\hat{m}_{\theta,e(i)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(i)} + \hat{m}_{P(\Theta),e(i)}} & \theta \neq \emptyset \end{cases} \tag{31}$$

$$\hat{m}_{\theta,e(i)} = \left[(1-w_i)m_{\theta,e(i-1)} + m_{P(\Theta),e(i-1)}m_{\theta,i}\right] + \sum_{B\cap C=\theta} m_{B,e(i-1)}m_{C,i}, \quad \forall\theta \subseteq \Theta \tag{32}$$

$$\hat{m}_{P(\Theta),e(i)} = (1-w_i)m_{P(\Theta),e(i-1)} \tag{33}$$

where w_i is the weight of e_i, which is not necessarily normalised, $0 \leqslant m_{\theta,e(i)}, m_{P(\Theta),e(i)} \leqslant 1$ and $\sum_{\theta\subseteq\Theta} m_{\theta,e(i)} + m_{P(\Theta),e(i)} = 1$ for $i = 2, \ldots, L$ recursively.

Proof. See Appendix A.4. □

Corollary 2 *(Recursive ER rule with weight). The combined degree of belief to which L pieces of independent evidence e_i with weights w_i $(i = 1, \ldots, L)$, which are not necessarily normalised, jointly support proposition θ is given by*

$$p_\theta = p_{\theta,e(L)} = \begin{cases} 0 & \theta = \emptyset \\ \frac{\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(L)}} & \theta \neq \emptyset \end{cases} \tag{34}$$

with $\hat{m}_{\theta,e(L)}$ given by Eqs. (32) for $i = L$, $0 \leqslant p_\theta \leqslant 1$ $\forall\theta \subseteq \Theta$ and $\sum_{\theta\subseteq\Theta} p_\theta = 1$.

Proof. See Appendix A.5. □

Corollary 1 is established to calculate the probability masses for the combined *WBD* of the first i pieces of evidence. It needs to be applied $L-1$ times in a recursive fashion. Corollary 2 is established to calculate the belief degrees of the final combined *BD* after all the L pieces of evidence are combined, and is applied only once at the end of the recursive process. Corollaries 1 and 2 constitute the *ER* rule for combining multiple pieces of independent evidence under the condition that each piece of evidence plays a limited role equal to its weight.

Table 1
Comparison of three rules for Case 1.

		$\emptyset$	A	B	Θ	$P(\Theta)$
Dempster's rule	m_1	0	0.8	0.2	0	0
	m_2	0	0.4	0.6	0	0
	$\sum_{C\cap D=\theta} m_1(C)m_2(D)$	0.56	0.32	0.12	0	0
	m	0	0.7273	0.2727	0	0
PCR5$_\emptyset$ rule	m_1	0	0.8	0.2	0	0
	m_2	0	0.4	0.6	0	0
	$\sum_{C\cap D=\theta} m_1(C)m_2(D)$	0.56	0.32	0.12	0	0
	m	0	0.6476	0.3524	0	0
ER rule	m_1	0	0.8	0.2	0	0
	m_2	0	0.4	0.6	0	0
	$m_{\theta,e(2)}$	0	0.7273	0.2727	0	0
	$p_{\theta,e(2)}$	0	0.7273	0.2727	0	0

3.6. First numerical comparison study

In this section, a simple example [28] is re-examined to illustrate how the following three rules deal with evidence weights differently: Dempster's rule (Eq. (5)) with Shafer's discounting method of Eq. (10), the *PCR5*$_\emptyset$ rule (Eq. (8) except that $m(\emptyset)$ is not forced to 0) with the *PCR5*$_\emptyset$ importance discounting method given by Eq. (11), and the *ER* rule with weight. The example is re-described using the notation of belief (probability) distribution defined by Eq. (20) as follows.

Example 1. Suppose $\Theta = \{A, B\}$ and there are two pieces of independent evidence e_1 and e_2 given by $e_1 = \{(A, 0.8), (B, 0.2)\}$ and $e_2 = \{(A, 0.4), (B, 0.6)\}$, both of which are complete without any ignorance, that is they are both Bayesian probability distributions.

Case 1 *(No importance discounting).* Suppose the two pieces of evidence are both very important and have the same maximal importance, or $w_1 = w_2 = 1$. The process and results generated using the three rules are shown in Tables 1. Note that in this case Eq. (29) reduces to Eq. (5), so the ***ER* rule reduces to Dempster's rule**. The results generated by *PCR5*$_\emptyset$ are different from those of Dempster's rule or the *ER* rule. This is not surprising as *PCR5*$_\emptyset$ is not a Bayesian process and adopts a relative weighting scheme to allocate the conflicting beliefs between A and B, so A and B are given different proportions of the conflicting beliefs. In Dempster' rule however, the conflicting beliefs earmarked to the empty set initially are re-distributed to the propositions A and B by the same proportion of (1–0.56).

Case 2 *(With importance discounting).* Now suppose the weights of e_1 and e_2 are different and are given for example by $w_1 = 0.2$ and $w_2 = 0.8$. Note that there is $w_1 + w_2 = 1$ in this case as the relative importance of e_1 and e_2 is taken into account. In general, this unity constraint does not have to be applied. The process and results generated using the three rules are shown in Tables 2. It is worth noting that as there is no ignorance in e_1 or e_2, the final result for $m(\Theta)$, the degree of global ignorance generated by Dempster's rule, as shown in the Θ column and the fifth row of Table 2, is incurred entirely due to using Shafer's discounting method (Eq. (10)), or assigning $m(P(\Theta))$ to Θ so $m(\Theta) = m(P(\Theta))$. If $m(P(\Theta))$ is then redistributed to A and B, or $p(A) = m(A)/(1 - m(\Theta)) = 0.3726/(1 - 0.1757) = 0.452$ and $p(B) = m(B)/(1 - m(\Theta)) = 0.4517/(1 - 0.1757) = 0.548$, these results will be the same as those generated by the *ER* rule, as shown in the last row of Table 2. If there is global ignorance in any evidence however, the separation between $m(\Theta)$ and $m(P(\Theta))$ and the redistribution of $m(P(\Theta))$ become impossible after the combination of evidence by Dempster's rule. Without the separation and redistribution, the final results shown in the fifth row of Table 2 are counter-intuitive as both pieces of evidence e_1 or e_2 are complete and Bayesian but their combined result becomes incomplete and non-Bayesian. This example shows that Shafer's discounting method leads to a non-Bayesian reasoning process. This also explains why the residual support $(1 - w_i)$ needs to be earmarked for or nominally allocated to $P(\Theta)$ rather than specifically assigned to any individual element of $P(\Theta)$ *a priori*.

In Table 2, the results generated by the *PCR5*$_\emptyset$ rule are close to those generated by the *ER* rule. However, this seems more coincidental for this particular example than general. This is because when applying *PCR5*$_\emptyset$ to get the *bpa* of a piece of evidence its specificity is changed, as beliefs are assigned to the empty set in *PCR5*$_\emptyset$. However, when applying the *ER* rule to get the *bpa* there is no change of specificity for any evidence. Also, the two rules forms different processes with *PCR5*$_\emptyset$ being a non-Bayesian process and the *ER* rule a conjunctive probabilistic reasoning process, or a generalised Bayesian reasoning process.

Table 2
Comparison of three rules for Case 2.

		$\emptyset$	A	B	Θ	$P(\Theta)$
Dempster's rule	m_1	0	0.16	0.04	0.8	0
	m_2	0	0.32	0.48	0.2	0
	$\sum_{C\cap D=\theta} m_1(C)m_2(D)$	0.0896	0.3392	0.4112	0.1600	0
	m	0	0.3726	0.4517	0.1757	0
PCR5$_\emptyset$ rule	m_1	0.8	0.16	0.04	0	0
	m_2	0.2	0.32	0.48	0	0
	$\sum_{C\cap D=\theta} m_1(C)m_2(D)$	0.9296	0.0512	0.0192	0	0
	m	0	0.43	0.57	0	0
ER rule	m_1	0	0.16	0.04	0	0.8
	m_2	0	0.32	0.48	0	0.2
	$m_{\theta,e(2)}$	0	0.3726	0.4517	0	0.1757
	$p_{\theta,e(2)}$	0	0.452	0.548	0	0

4. New *ER* rule to combine evidence with both weight and reliability

Shafer's discounting [25] has been widely regarded as reliability discounting [14]. As discussed in the introduction section, there is clear difference between evidence importance and evidence reliability. In information fusion, for example, the importance of an information source is indicated by weight as granted to the source by a fusion system designer, which can be relative to other information sources; the reliability of an information source represents its ability to provide correct assessment or solution for a given problem [28]. In other words, the reliability and importance of a piece of evidence may not measure the same property of the evidence, and as such they need to be treated separately in an inference process. In this section, we first review how this was handled in the enhanced *PCR5* rule, which seems to be the closest to what is established in this paper. We will then enhance the *ER* rule to deal with both evidence importance and reliability in a unified and coherent framework of weighted belief distributions with reliability, followed by a comparison study.

4.1. Introduction to the PCR5 rule to combine evidence with weight and reliability

In the enhanced *PCR5* rule, weight and reliability are taken into account by using Shafer's discounting method to deal with reliability and the *PCR5* importance discounting method (Definition 2) to handle weight [28]. The basic idea is to generate *bpa* by first discounting the ith piece of evidence with its reliability using Shafer's discounting method (Eq. (10)), followed by discounting the discounted evidence with its weight using the *PCR5* importance discounting method (Eq. (11)), with the generated *bpa* represented by m_{r_i,w_i}. An alternative way is to change the order of the two types of discounting, with the generated *bpa* represented by m_{w_i,r_i}. In general, it is the case that $m_{r_i,w_i} \neq m_{w_i,r_i}$, and therefore different methods could be developed to combine these two types of *bpa*. Since these methods are fairly similar, we only introduce the first method.

***PCR5* Method 1.** This method consists of the following steps:

Step 1: Apply Shafer's discounting method to discount evidence by reliability and then the *PCR5* importance discounting method to further discount the discounted evidence by weight, leading to $m_{r_i,w_i}(\theta)$ for all θ and $i = 1,\dots,L$. Then combine $m_{r_i,w_i}(\theta)$ for all $i = 1,\dots,L$ for each θ using *PCR5*$_\emptyset$, and normalise the combined *bpa*, which is represented by $\bar{m}_{PCR5\emptyset,r,w}(\theta)$.

Step 2: Reverse the order of the first part in Step 1 to get $m_{w_i,r_i}(\theta)$ for all θ and $i = 1,\dots,L$. Then repeat the next two parts of Step 1 to get $\bar{m}_{PCR5\emptyset,w,r}(\theta)$.

Step 3: Combine the *bpa*s generated from Step 1 and Step 2 using some combination rule such as the arithmetic mean operator, to get the final result, given by $\bar{m}_{PCR5}(\theta)$.

The above *PCR5* method is straightforward and relatively simple if any possible complexity of the *PCR5* rule is taken out of the picture. Since the two discounting methods each change the specificity of evidence as mentioned above, the process of this method is no longer Bayesian after the implementation of Shafer's discounting method for reliability, nor is it a probable reasoning process as defined in *D–S* theory after the implementation of the *PCR5* importance discounting method for weight. This reasoning process may be interpreted as a plausible and paradoxical reasoning process as for the *PCR5* rule. So it remains as a question how to interpret the results generated by using the *PCR5* Method 1.

4.2. New ER rule to combine evidence with weight and reliability

In Section 3.3, the weight of a piece of evidence was used to define *WBD* for representing the degree of support to each proposition, with the residual support, measured by $1 - w_i$, earmarked to the power set. This approach retains the specificity of the original evidence as represented by a belief distribution in the processes of both generating *bpa* and combining support by the orthogonal sum operation. As such, it is a conjunctive probabilistic reasoning process or a Bayesian process when used to combine probability distributions.

The role of the residual support can be explained by examining the *ER* rule. As shown in Eq. (30), the real purpose of the residual support $1-w_1$ (or $1-w_2$) is to set a bound within which the other piece of evidence can play a limited role. For example, if $1-w_1=0$, there will be $(1-w_1)m_{\theta,2}=0$. This means that if $1-w_1=0$ the individual support for θ from the second piece of evidence e_2 will not be counted at all, however reliable e_2 may be. At the other extreme, if $1-w_1=1$, the individual support from e_2 will be fully counted. The same can be said for the first piece of evidence if $1-w_2=0$ or $1-w_2=1$. This shows that the residual support $1-w_i$ acts as the unreliability of the ith piece of evidence. The above analysis leads to the following more generic definition.

Definition 4 *(ER weighting with reliability).* Suppose w_i is the weight of evidence e_i defined in Eq. (20) with $0 \leqslant w_i \leqslant 1$ and r_i is the reliability of e_i with $0 \leqslant r_i \leqslant 1$ and $r_i=0$ and 1 standing for "not reliable at all" and "fully reliable" respectively. The basic probability masses for e_i are then assigned as follows

$$\tilde{m}_{\theta,i} = \begin{cases} 0 & \theta=\emptyset \\ c_{rw,i}m_{\theta,i} & \theta \subseteq \Theta, \theta \neq \emptyset \\ c_{rw,i}(1-r_i) & \theta = P(\Theta) \end{cases} \tag{35}$$

$$c_{rw,i} = 1/(1+w_i-r_i) \tag{36}$$

In Definition 4, $\tilde{m}_{\theta,i}$ measures the degree of support for θ from e_i with both the weight and reliability of e_i taken into account. $1-r_i$ is the unreliability of evidence e_i, which sets a bound on the role that other evidence can play when combined with e_i. In particular, there is $\tilde{m}_{P(\Theta),i}=0$ when $r_i=1$, so any θ will be ruled out by e_i if $\tilde{m}_{B,i}=0$ for any $B \cap \theta = \theta$, whatever support θ may get from other evidence. $c_{rw,i}$ is a normalisation factor such that $\sum_{\theta\subseteq\Theta}\tilde{m}_{\theta,i}+\tilde{m}_{P(\Theta),i}=1$ given $m_{\theta,i}=w_i p_{\theta,i}$ and $\sum_{\theta\subseteq\Theta}p_{\theta,i}=1$.

Let $\tilde{w}_i = c_{rw,i}w_i$. In Eqs. (35), $c_{rw,i}m_{\theta,i}$ is then equal to $\tilde{w}_i p_{\theta,i}$ and $c_{rw,i}(1-r_i)$ equal to $1-\tilde{w}_i$. As such, Eq. (35) can be equivalently rewritten as follows

$$\tilde{m}_{\theta,i} = \begin{cases} 0 & \theta=\emptyset \\ \tilde{w}_i p_{\theta,i} & \theta \subseteq \Theta, \theta \neq \emptyset \\ 1-\tilde{w}_i & \theta = P(\Theta) \end{cases} \tag{37}$$

which is consistent with Definition 3 with $\tilde{w}_i$ acting as a new weight. Hence, $c_{rw,i}$ also acts as a regulative coefficient to adjust w_i according to reliability r_i, resulting in a hybrid weight $\tilde{w}_i$ with $\tilde{w}_i < w_i$ ($c_{rw,i}<1$) if $r_i<w_i$, $\tilde{w}_i=w_i$ ($c_{rw,i}=1$) if $r_i=w_i$, and $\tilde{w}_i > w_i$ ($c_{rw,i}>1$) if $r_i>w_i$. Furthermore, if e_i is fully reliable, or $r_i=1$, there will be $\tilde{w}_i=1$; if e_i is not fully reliable, or $r_i<1$, there will be $\tilde{w}_i<1$. Since $c_{rw,i}\geqslant 0$ and $w_i \geqslant 0$, there is always $\tilde{w}_i \geqslant 0$. $\tilde{w}_i$ can thus be interpreted as a hybrid weight and reliability coefficient for e_i, or the adjusted weight of e_i, to measure the degree of support from e_i with $1 \geqslant \tilde{w}_i \geqslant 0$. From Definition 4, a piece of evidence can be represented by

$$m_i = \{(\theta, \tilde{m}_{\theta,i}),\ \forall\theta\subseteq\Theta;\ \big(P(\Theta), \tilde{m}_{P(\Theta),i}\big)\} \tag{38}$$

The above distribution is called *Weighted Belief Distribution with Reliability* (*WBDR*).

It is worth noting that in Definition 4 the original specificity of the belief distribution of e_i as represented by $p_{\theta,i}$ is strictly preserved because the following equation always holds

$$p_{\theta,i} = \frac{\tilde{m}_{\theta,i}}{1-\tilde{m}_{P(\Theta),i}}, \quad \forall\theta\subseteq\Theta \tag{39}$$

Given the above discussions, the new *ER* rule with weight and reliability taken into account results from the orthogonal sum of two *WBDR*s as follows.

Corollary 3 *(ER rule with weight and reliability). Suppose two pieces of independent evidence e_1 and e_2 are each profiled by Eq. (20) and their WBDRs are represented by Eq. (38). The combined degrees of belief to which e_1 and e_2 jointly support proposition θ, denoted by $p_{\theta,e(2)}$, is given by Eq. (29) with $\hat{m}_{\theta,e(2)}$ given by*

$$\hat{m}_{\theta,e(2)} = \big[(1-r_2)m_{\theta,1}+(1-r_1)m_{\theta,2}\big] + \sum_{B\cap C=\theta} m_{B,1}m_{C,2} \quad \forall\theta\subseteq\Theta \tag{40}$$

Proof. Note that $\tilde{m}_{\theta,i}=c_{rw,i}m_{\theta,i}$ for any $\theta\subseteq\Theta$ and $\tilde{m}_{P(\Theta),i}=c_{rw,i}(1-r_i)$ for $i=1,2$. The proof is then the same as for proving Theorem 2 by implementing the orthogonal sum operation on the *WBDR*s of e_1 and e_2 with $c_{rw,i}$ cancelled out in the normalisation. □

To combine multiple pieces of evidence, we have the following results.

Table 3
Three pieces of independent evidence.

$p_{\theta,i}$	$\emptyset$	A	B	C	$\{A, B\}$	$\{A, C\}$	$\{B, C\}$	$\{A, B, C\}$
e_1	0	0.8	0	0	0.1	0.1	0	0
e_2	0	0.4	0.3	0	0.2	0	0.1	0
e_3	0	0.1	0.3	0.5	0	0	0	0.1

Corollary 4 *(Recursive ER rule with weight and reliability). Suppose L pieces of independent evidence are each profiled by Eq. (20) and their WBDRs are represented by Eq. (38). Suppose $e(i)$ is defined in the same way as in Corollary 1. The combined degree of belief to which L pieces of independent evidence e_i with weight w_i and reliability r_i ($i = 1, \ldots, L$) jointly support proposition θ is given by Eq. (34) with $\hat{m}_{\theta,e(L)}$ generated by recursively applying Eq. (31) and the following two equations*

$$\hat{m}_{\theta,e(i)} = \left[(1 - r_i)m_{\theta,e(i-1)} + m_{P(\Theta),e(i-1)}m_{\theta,i}\right] + \sum_{B \cap C = \theta} m_{B,e(i-1)}m_{C,i}, \quad \forall \theta \subseteq \Theta \tag{41}$$

$$\hat{m}_{P(\Theta),e(i)} = (1 - r_i)m_{P(\Theta),e(i-1)} \tag{42}$$

Proof. Note that $\tilde{m}_{\theta,i} = c_{rw,i}m_{\theta,i}$ for any $\theta \subseteq \Theta$ and $\tilde{m}_{P(\Theta),i} = c_{rw,i}(1 - r_i)$ for $i = 1, \ldots, L$. The proof is then the same as for proving Corollaries 1 and 2 by implementing the orthogonal sum operation recursively on the *WBDR*s of e_i for $i = 1, \ldots, L$, followed by implementing Eq. (34) once at the end with $c_{rw,i}$ cancelled out in the normalisation. $\square$

It should be noted that in both Corollaries 3 and 4 the orthogonal sum operation was used to combine *WBDR*s. As such, the above *ER* rule with evidence weight and reliability constitutes a conjunctive probabilistic reasoning process and is a Bayesian inference process when used to combine probability information.

We can now calculate overall reliability. Let $r_{e(L)}$ stand for the combined reliability of all L pieces of evidence $e(L)$ and $w_{e(L)}$ for the combined weight of $e(L)$. From Eqs. (36) and (37), we get $m_{P(\Theta),e(L)} = (1 - r_{e(L)})/(1 + w_{e(L)} - r_{e(L)})$. With $m_{P(\Theta),e(L)}$ generated by Corollary 4 and $w_{e(L)}$ judged by the decision maker, we can get $r_{e(L)}$ as follows

$$r_{e(L)}(w_{e(L)}) = \left(1 - m_{P(\Theta),e(L)}(1 + w_{e(L)})\right)/(1 - m_{P(\Theta),e(L)}) \tag{43}$$

It should be noted that the above equation does not define the relationship between reliability and weight. Rather, it is simply deduced to calculate the combined reliability from the generated residual support $m_{P(\Theta),e(L)}$, which depends on how $w_{e(L)}$ is judged. If $w_{e(L)} = r_{e(L)}$, there will be $r_{e(L)} = 1 - m_{P(\Theta),e(L)}$. If a precise judgement for $w_{e(L)}$ cannot be provided, the range of possible values for $w_{e(L)}$ can be decided by noting that the maximal value for $w_{e(L)}$ is 1 and the minimal value of $w_{e(L)}$ should not be smaller than the maximum of the weights of the L pieces of evidence combined. So, in general $r_{e(L)}$ is given by

$$r_{e(L)}(w_{e(L)}) \in \left[r_{e(L)}(1), r_{e(L)}\left(\max\{w_i\}\right)\right]. \tag{44}$$

It is worth noting from Corollary 4 that $m_{P(\Theta),e(L)}$ will be zero if and only if any piece of evidence is fully reliable. From Eq. (43), it can then be said that $r_{e(L)}$ will be one if and only if any of the L pieces of evidence is fully reliable, and otherwise less than one.

4.3. Second numerical comparison study

From the literature, only the *PCR5* Method as summarised in Section 4.1 is found comparable with the new *ER* rule in terms of distinctively taking into account both the weight and the reliability of evidence in combination. In this section, these two rules are compared using the same example as examined in the literature [28].

Suppose $\Theta = \{A, B, C\}$ with A, B and C mutually exclusive and collectively exhaustive, and three pieces of independent evidence e_1, e_2 and e_3 are represented by three belief distributions as shown by the 2nd to 4th rows of Table 3. Suppose their reliabilities are given by $r_1 = 0.8$, $r_2 = 0.5$ and $r_3 = 0.2$ respectively, and their weights by $w_1 = 0.9$, $w_2 = 0.3$ and $w_3 = 0.6$ respectively.

The results generated by using the *PCR5* Method 1 were originally provided in Tables 5 to Table 7 of the paper authored by Smarandache et al. [28]. These results are rearranged in a consistent way with the other results of this paper, as shown in Table 4. In rows 2–4 of Table 4 headed by $m_{r_1,w_1}(\theta)$, $m_{r_2,w_2}(\theta)$ and $m_{r_3,w_3}(\theta)$, the reliability-importance discounting is performed for e_1, e_2 and e_3 by applying Shafer's reliability discounting method first (Eq. (10) with α replaced by r_i), followed by applying the *PCR5* importance discounting method (Eq. (11)). Similarly, the importance-reliability discounting (Eq. (11) applied first and then Eq. (10) with α replaced by r_i) is performed as shown in rows 5–7 of Table 4 headed by $m_{w_1,r_1}(\theta)$, $m_{w_2,r_2}(\theta)$ and $m_{w_3,r_3}(\theta)$. The results of Steps 1, 2 and 3 of the *PCR5* Method 1 are given by rows 8, 9 and 10 of Table 4 headed by $\bar{m}_{PCR5\emptyset,r,w}(\theta)$, $\bar{m}_{PCR5\emptyset,w,r}(\theta)$ and $\bar{m}_{PCR5}(\theta)$ respectively.

The results generated by using the *ER* rule are shown in Table 5. The support for each proposition from every piece of evidence is given as *bpa* generated by using Definition 4, as shown in rows 2–4 of Table 5 headed by $\tilde{m}_{\theta,1}$, $\tilde{m}_{\theta,2}$ and $\tilde{m}_{\theta,3}$

Table 4
Results generated by the *PCR5* Method.

	$\emptyset$	A	B	C	$\{A,B\}$	$\{A,C\}$	$\{B,C\}$	$\{A,B,C\}$
$m_{r_1,w_1}(\theta)$	0.1	0.576	0	0	0.072	0.072	0	0.18
$m_{r_2,w_2}(\theta)$	0.7	0.06	0.045	0	0.03	0	0.015	0.15
$m_{r_3,w_3}(\theta)$	0.4	0.012	0.036	0.06	0	0	0	0.492
$m_{w_1,r_1}(\theta)$	0.08	0.576	0	0	0.072	0.072	0	0.2
$m_{w_2,r_2}(\theta)$	0.35	0.06	0.045	0	0.03	0	0.015	0.5
$m_{w_3,r_3}(\theta)$	0.08	0.012	0.036	0.06	0	0	0	0.812
$\bar{m}_{PCR5\emptyset,r,w}(\theta)$	0	0.5741	0.0254	0.0182	0.0311	0.0233	0.0032	0.3247
$\bar{m}_{PCR5\emptyset,w,r}(\theta)$	0	0.4927	0.0244	0.0182	0.0464	0.0386	0.0032	0.3765
$\bar{m}_{PCR5}(\theta)$	0	0.5334	0.0249	0.0182	0.0388	0.0310	0.0032	0.3506

Table 5
Results generated by the *ER* rule.

	$\emptyset$	A	B	C	$\{A,B\}$	$\{A,C\}$	$\{B,C\}$	$\{A,B,C\}$	$P(\Theta)$
$\tilde{m}_{\theta,1}$	0	0.6545	0.0000	0.0000	0.0818	0.0818	0.0000	0.0000	0.1818
$\tilde{m}_{\theta,2}$	0	0.1500	0.1125	0.0000	0.0750	0.0000	0.0375	0.0000	0.6250
$\tilde{m}_{\theta,3}$	0	0.0429	0.1286	0.2143	0.0000	0.0000	0.0000	0.0429	0.5714
$\hat{m}_{\theta,e(2)}$	0	0.5406	0.0288	0.0027	0.0624	0.0450	0.0060	0.0000	0.1000
$m_{\theta,e(2)}$	0	0.6882	0.0367	0.0034	0.0794	0.0573	0.0076	0.0000	0.1273
$\hat{m}_{\theta,e(3)}$	0	0.6490	0.0767	0.0617	0.0683	0.0493	0.0066	0.0076	0.1018
$m_{\theta,e(3)}$	0	0.6356	0.0751	0.0604	0.0669	0.0483	0.0064	0.0075	0.0997
$p_{\theta,e(3)}$	0	0.7061	0.0835	0.0671	0.0743	0.0536	0.0071	0.0083	

respectively. $\hat{m}_{\theta,e(2)}$ and $m_{\theta,e(2)}$, followed by $\hat{m}_{\theta,e(3)}$ and $m_{\theta,e(3)}$, in rows 5–8 of Table 5 are generated by applying Corollary 3 recursively, and $p_{\theta,e(3)}$ in row 9 of Table 5 by using Corollary 4.

The final results of the *PCR*5 method shown in the last row of Table 4 are significantly different from those of the *ER* rule shown in the last row of Table 5. A question arises as to which results could be trusted and if so why? Comparing the three pieces of evidence given in rows 2–4 of Table 3 with those of rows 2–4 of Table 4 generated by the reliability-importance discounting, it can be noted that the specificity of the original evidence is changed. For example, e_1 does not contain any global ignorance with $p_{\{A,B,C\},1} = 0$ as shown in the 2nd row and last column of Table 3, but after the discounting there is $m_{r_1,w_1}(\{A,B,C\}) = 0.18$ as shown in the 2nd row and last column of Table 4. There are similar specificity changes for e_2 and e_3. For e_3, the specificity change leads to difficulty in interpreting *bpa* and consequently the final results. This is because e_3 does include global ignorance with $p_{\{A,B,C\},3} = 0.1$, as shown in the last row and last column of Table 3 which is then mixed with the residual support of e_3, or $1 - w_3$, in the discounting and/or combination processes. The outcome $m_{r_3,w_3}(\{A,B,C\}) = 0.492$ in the 4th row and last column of Table 4 is a mix of two different things: global ignorance and residual support. The discounting also leads to basic probability masses being given to the empty set for e_1, e_2 and e_3 with $m_{r_1,w_1}(\emptyset) = 0.1$, $m_{r_2,w_2}(\emptyset) = 0.7$ and $m_{r_3,w_3}(\emptyset) = 0.4$ as shown in rows 2–4 and column 2 of Table 4. One of the consequences of these changes of specificity is that the three distributions shown in rows 2–4 of Table 4 are no longer a belief distribution as defined in *D–S* theory and that the numbers in the distributions need to be reinterpreted because they are not probability as defined in probability theory or belief as defined in *D–S* theory. There are similar changes of specificity after the importance-reliability discounting method is employed, as shown in rows 5–7 of Table 4.

As discussed before, the *PCR5* rule is not a probabilistic reasoning process, so the result it generates is not a probability, nothing to say that the specificity of each piece of evidence has already been changed even before the *PCR5* rule is applied. As such, there is a question as to how to interpret what the results generated by *PCR5* actually mean. By looking at the final result shown in the last row of Table 4, one obvious concern is that a large portion of the final result is generated as global ignorance with $\bar{m}_{PCR5}(\{A,B,C\}) = 0.3506$, as shown in the last row and last column of Table 4. Hence it may be concluded from this final result that:

- it is most plausible that proposition A is true with a plausibility degree of at least 53.34%, which could be as high as 95.38%, or $(0.5334 + 0.0388 + 0.031 + 0.3506) \times 100$, and
- any other proposition could also be true with a range of plausibility degrees that could be as high as over 35.06%.

The question is whether such conclusions make sense or not. To answer this question, it is necessary to look at the original data and use intuitive reasoning qualitatively. It is obvious from Table 3 that proposition A should be the most likely to occur among all propositions. So the above first conclusion does make sense. However, is the range of plausibility degrees [53.34%, 95.38%] appropriate? To answer this question, let's examine the original and generated data for $\{A,C\}$, $\{B,C\}$ and $\{A,B,C\}$. In the original data, each of them is only given a belief degree of 0.1 in e_1, e_2 and e_3 respectively, or $p_{\{A,C\},1} = 0.1$, $p_{\{B,C\},2} = 0.1$ and $p_{\{A,B,C\},3} = 0.1$ as shown in Table 3. It thus seems reasonable to expect that none of them should be given a larger degree of belief than 0.1 in the final results after the three pieces of evidence are combined. In the final result shown in the last row of Table 4, there are $\bar{m}_{PCR5}(\{A,C\}) = 0.031$ and $\bar{m}_{PCR5}(\{B,C\}) = 0.0032$, which look reasonable. However, we have $\bar{m}_{PCR5}(\{A,B,C\}) = 0.3506$, which is much larger than 0.1, inconsistent with the expectation.

A question then arises as to whether it is appropriate to conclude that one would still be 35.06% totally ignorant about which proposition or subset of propositions is true after combining the three pieces of evidence. In the three original pieces of evidence, only one of them contains a small portion (10%) of total ignorance, which nevertheless is the least reliable by far and is not that important either. A commonsense answer to this question seems to be no, leading to doubt over the accuracy of the *PCR*5 results. In other words, the range of plausibility degrees between 53.34% and 95.38% may exaggerate the ambiguity of the probability to which proposition A is true, even though this range of plausibility might embrace the true range of the probability. Such exaggeration can also be found for other propositions, which could be misleading. For example, it does not look rational to conclude that proposition $\{B,C\}$ could be true with a plausibility degree range of 0.0463 $(0.0249+0.0182+0.0032)$ to as high as 0.4667 $(0.0249+0.0182+0.0388+0.031+0.0032+0.3506)$, since only e_2 points to it with a belief degree of 0.1 and e_3 only potentially points to it by a total degree of 0.1, both of which are much less important or reliable than e_1.

The results shown in Table 5 are straightforward to interpret as the *ER* rule constitutes a conjunctive probabilistic reasoning process. In fact, the *bpa*s given in rows 2–4 of Table 5 do not change the specificity of any evidence at all. For example, the *bpa* for e_1, as shown in row 2 of Table 5, was generated for all $\theta \subseteq \Theta$ by multiplying the belief degrees shown in row 2 of Table 3 by the same coefficient $c_{rw,1}w_1 = 0.8182$ with $r_1 = 0.8$, $w_1 = 0.9$ and $c_{rw,1} = 1/(1+w_1-r_1) = 0.9091$. The residual support for e_1 is assigned to the power set, given by $m_{P(\Theta),1} = c_{rw,1}(1-r_1) = 0.9091 \times (1-0.8) = 0.1818$, shown in the 2nd row and last column of Table 5. The sum of all elements in this row is one and the corresponding distribution is a belief distribution. The combined *bpa* shown by $m_{\theta,e(2)}$ was generated by combining the *bpa* for e_1 and the *bpa* for e_2 by means of the orthogonal sum of the two *WBDR*s characterised by $\tilde{m}_{\theta,1}$ and $\tilde{m}_{\theta,2}$. $m_{\theta,e(2)}$ is then combined with $\tilde{m}_{\theta,3}$ to generate $m_{\theta,e(3)}$, which is then normalised to finally get $p_{\theta,e(3)}$. So the whole process of combining the three pieces of evidence is conjunctive. From this conjunctive probabilistic reasoning process, by looking at the figures in the last row of Table 5 and using Eq. (3), it can be concluded that

- Proposition A is the most likely to occur with a high probability of 0.7061 to 0.8423;
- Proposition B is less likely to occur with a low probability of 0.0835 to 0.1732;
- Proposition C is also less likely to occur with a low probability of 0.0671 to 0.1361;
- Proposition $\{A,B\}$ has the high degrees of belief and plausibility of 0.8639 and 0.9329 respectively, mainly due to the high probability to which proposition A may occur;
- Proposition $\{A,C\}$ has the high degrees of belief and plausibility of 0.8268 and 0.9165, also mainly due to the high probability to which proposition A may occur;
- Proposition $\{B,C\}$ has the low degrees of belief and plausibility of 0.1577 and 0.2939;
- The degree of the combined global ignorance for proposition $\Theta = \{A,B,C\}$ is as small as 0.0083, far less than 0.1 as observed in e_3 only, which is significantly lessened by the other two pieces of more important evidence that do not have any global ignorance;
- Finally the reliability of the above results of the *ER* rule ranges from 0.8892 to 0.9003, generated by using Eqs. (43) and (44).

The above conclusions seem agreeable with the raw information contained in the three original pieces of evidence. The final belief distribution generated by the *ER* rule provides a panoramic view on the probabilities of the occurrence of propositions, which can lay a solid foundation for further analysis and decision making.

5. Properties of the new *ER* rule

As a conjunctive probabilistic reasoning process, the new *ER* rule holds a number of properties. In this section, we first analyse its inherent properties in order to facilitate its appropriate use. This is followed by showing that both the *ER* algorithm and Dempster's rule are the special cases of the *ER* rule. We will then identify the missing definition of Dempster's rule when pieces of evidence are completely conflicting, and finally show how to combine highly conflicting evidence through a so-called reliability perturbation analysis.

5.1. Basic properties of the ER rule

The *ER* rule is based on the orthogonal sum operated on *WBD*s (*WBDR*s). As such, it inherits the basic properties of the orthogonal sum operator and is both commutative and associative. Note however that these properties prevail only if *WBD*s (*WBDR*s) are combined recursively until all of them are combined before generating the overall combined degree of belief. These properties ensure that the *ER* rule can be used to combine multiple pieces of evidence in any order without changing the final result.

From Eqs. (30), (32), (40) and (41), it is clear that if a piece of evidence is given a zero weight by a decision maker then the evidence will not play any role in the process of evidence combination. This means that adding a piece of evidence with zero weight into a list of evidence does not affect the combination result, so evidence with zero weight can be regarded as a neutral element. Note that the above property does not hold for a piece of evidence with zero reliability unless the evidence is subsequently given zero weight.

However, the combination of two identical pieces of evidence using the *ER* rule does not normally produce the same evidence unless they each point to propositions equally. This is because the *ER* rule is a conjunctive rule and accumulates joint support. So a proposition with a high (low) degree of support from each identical piece of evidence will be reinforced (weakened) more than proportionally. Hence, the *ER* rule is not an idempotent process.

Apart from the above basic properties, the *ER* rule also satisfies synthesis axioms [41] which any rational reasoning process ought to follow. While these axioms should hold for a rigorous combination algorithm, an initial study suggested that human intuition or System 1 thinking [18] may not always follow such axioms [12]. Such violation may be due to various biases and irrationality in human judgements as studied and reported widely [1]. The accuracy of mental information processing under time pressure and participants' lack of serious responsibility for the accuracy may also play a part in the reported violation of the axioms. The *ER* rule is developed for rational and rigorous reasoning but not as a descriptive model to explain human judgements. Its rigorous rationality check is thus the focus of this subsection.

We first restate the four synthesis axioms introduced previously [41] in the context of *BD*s and then prove that the *ER* rule provides a rational reasoning process for combining multiple pieces of independent evidence, in the sense that it satisfies the axioms.

Axiom 1 *(No support).* There should be no combined support for a proposition from multiple pieces of evidence if none of them supports the proposition at all.

Axiom 2 *(Consensus).* There should be full (100%) combined support for a proposition if all pieces of evidence each fully support the proposition.

Axiom 3 *(Locality).* There should be full combined support for a set and its subsets of propositions, but no combined support for any other proposition if all pieces of evidence each fully support this same set and its subsets of propositions only.

Axiom 4 *(Non-dominance).* There should be some combined support for a proposition if at least one piece of evidence supports the proposition and no evidence is dominant.

In the rest of this subsection, we show that the *ER* rule satisfies all the above four synthesis axioms by proving the following four theorems. Suppose there are L pieces of independent evidence described by *BD*s, as defined by Eq. (20). Let w_i and r_i be the weight and reliability of the ith piece of evidence which satisfies the following conditions

$$0 < w_i < 1 \quad \text{and} \quad 0 < r_i < 1 \quad \text{for any } i = 1, \dots, L \tag{45}$$

Suppose each piece of evidence plays a limited role and the combination of the L pieces of evidence is generated by using the *ER* rule given by Theorem 2, Corollaries 1 and 2, or Corollaries 3 and 4. Let their combined *BD* be described by

$$e(L) = \big\{(\theta, p_\theta),\ \forall\theta \subseteq \Theta\big\} \tag{46}$$

with p_θ given by Eq. (34). Then we have the following results.

Theorem 3 *(No support). The ER rule satisfies Axiom 1, that is there will be $p_\theta = 0$ if $p_{B,i} = 0$ for all $i = 1, \dots, L$ and any $B \subseteq \Theta$ with $B \cap \theta = \theta$.*

Proof. See Appendix A.6. □

Theorem 4 *(Consensus). The ER rule satisfies Axiom 2, that is if $p_{\theta,i} = 1$ and $p_{B,i} = 0$ for all $i = 1, \dots, L$ and any $B \subseteq \Theta$ with $B \neq \theta$, there will be $p_\theta = 1$ and $p_B = 0$ for any $B \subseteq \Theta$ with $B \neq \theta$.*

Proof. See Appendix A.7. □

Theorem 5 *(Locality). The ER rule satisfies Axiom 3, that is if all focal elements of the L BDs and their subsets are denoted by $\Theta^+ = \{B|B \cap \theta = B$ for any $p_{\theta,i} > 0,\ \theta \subseteq \Theta,\ i = 1, \dots, L\}$, there will be $p_B = 0$ for any $B \notin \Theta^+$ and $\sum_{\theta\in\Theta^+} p_\theta = 1$.*

Proof. See Appendix A.8. □

Theorem 6 *(Non-dominance). The ER rule satisfies Axiom 4, that is if there is $p_{\theta,i} > 0$ for at least one $i \in \{1, \dots, L\}$, there will be $p_\theta > 0$.*

Proof. See Appendix A.9. □

5.2. Relationship with the ER algorithm

The *ER* algorithm was developed for multiple criteria decision analysis in particular [41]. The *ER* rule is established primarily to generalise the *ER* algorithm for combination of evidence in general. It is therefore natural to conjecture that the *ER* algorithm should be a special case of the *ER* rule. This is indeed implied in the discussions in the previous sections. To formalise their relationship, the following corollary is provided.

Corollary 5 *(The ER algorithm). The ER algorithm is a special case of the ER rule when the reliability of each piece of evidence is equal to its weight which is normalised to be relative to each other as defined by Eq. (13).*

Proof. This is self-evident by normalising w_i so that $\sum_{i=1}^{L} w_i = 1$ and letting $r_i = w_i$ for all $i = 1, \ldots, L$. Putting the normalised w_i and r_i into Definition 4 leads to $c_{rw,i} = 1/(1 + w_i - r_i) = 1$, so Definition 4 reduces to Definition 3 and the conclusion results from the proofs of Lemma 1, Theorems 1 and 2. $\square$

5.3. Relationship with Dempster's rule

Both Dempster's rule and the *ER* rule originate from the orthogonal sum with the former operated on *BD*s and the latter on *WBD*s (or *WBDR*s). It is therefore natural that they should be closely related to each other. This is indeed the case, and their relationship can be formalised by the following corollary.

Corollary 6 *(Dempster's rule). Dempster's rule is a special case of the ER rule when all pieces of evidence are each fully reliable, or $r_i = 1$ for all $i = 1, \ldots, L$.*

Proof. See Appendix A.10. $\square$

Corollary 6 provides a basis for identifying the missing definition in Dempster's rule when it is used to combine pieces of evidence that are highly or completely conflicting. This can be achieved by exploring the *ER* rule through the analysis to be investigated in the rest of this section. We name it reliability perturbation analysis. The main idea can be traced back to the interpretation of conflict by Dubois and Prade [11] that if conflict between two information sources is high at least one of them may be unreliable. This interpretation was used to develop robust combination rules [14]. In this section, high conflict is treated as doubt over the reliability of information sources. The doubt is explored by means of perturbing the reliability of evidence as follows.

Let σ denote the total degree of collective support from two pieces of independent evidence e_1 and e_2 that each are assumed to be fully reliable and have the weight of 1, or

$$\sigma = \sum_{B \cap C = D,\, D \subseteq \Theta} p_{B,1} p_{C,2} \tag{47}$$

A small σ thus means a high level of conflict between e_1 and e_2. Suppose τ is a small real number that means "unlikely" cognitively such as $\tau = 0.01$.

If e_1 and e_2 are each fully reliable and their collective support to the same proposition is deemed to be cognitively likely, i.e. $\sigma > \tau$, Dempster's rule can be applied to combine their *BD*s. If they are completely conflicting, i.e. $\sigma = 0$, it is conjectured that the weighted average rule should be applied to combine their *BD*s. This conjecture is based on the observation that the conjunction of e_1 and e_2 will be empty if they are completely conflicting, which is a special case that e_1 and e_2 are mutually exclusive. It is widely accepted that the combination of two pieces of mutually exclusive evidence is additive.

If $\tau \geqslant \sigma > 0$, e_1 and e_2 are deemed to be highly conflicting. In this case, to reflect the above interpretation of conflict, the reliabilities of e_1 and e_2 (r_i for $i = 1, 2$) are perturbed (or reduced slightly) to cast doubt over their reliabilities rather than assume that r_i has to become very small or zero simply because e_1 and e_2 are in high conflict. For example, the following reliability perturbation function can be used to analyse conflict resolution,

$$r_i(\sigma) = \begin{cases} 1 - \frac{b_i(\tau^k - \sigma^k)}{1 - \tau^k} & \text{if } \tau^2 \leqslant \sigma \leqslant \tau \\ 1 - b_i \sigma^{k/2} & \text{if } 0 \leqslant \sigma \leqslant \tau^2 \end{cases} \quad \text{with } \tau \leqslant b_i \leqslant 1/\tau \text{ and } 0 < k \leqslant 1 \tag{48}$$

In Eq. (48), b_i is referred to as a reliability perturbation coefficient, describing the level of the *Decision Maker*'s (*DM*'s) doubt over the reliability of the ith piece of evidence in light of its high conflict with other evidence and also the *DM*'s attitudes towards individual and collective support. A large b_i means a big reduction in reliability r_i. k is introduced into the equation to ensure that when $\sigma \to 0$ or $\sigma \to \tau$, $r_i(\sigma)$ will approach one not faster than σ approaching zero or τ. Note that we have $r_i(\sigma) = 1$ at $\sigma = 0$ or $\sigma = \tau$ and $0 \leqslant r_i(\sigma) < 1$ when $0 < \sigma < \tau$. Also note that $r_i(\sigma)$ is the smallest at $\sigma = \tau^2$. In other words, the maximal level of doubt over the reliability of e_i is $b_i \tau^k$ which can happen when e_1 and e_2 become unlikely to support a proposition collectively. It should be noted that what reliability perturbation function should

Table 6
Two pieces of independent evidence in high conflict.

$p_{\theta,i}$	$\emptyset$	A	B	C	$\{A,B\}$	$\{A,C\}$	$\{B,C\}$	$\{A,B,C\}$
e_1	0	$1-\delta$	δ	0	0	0	0	0
e_2	0	0	δ	$1-\delta$	0	0	0	0

be used is problem specific. As a rule of thumb, a reliability perturbation function should be a continuous function having the following properties: (1) $0 \leqslant r_i(\sigma) \leqslant 1$; (2) $r_i(\sigma) = 1$ at $\sigma = 0$ or $\sigma = \tau$; (3) when $\sigma \to 0$ or $\sigma \to \tau$, there should be $r_i(\sigma) \to 1$ at a speed no more than σ approaching zero or τ.

For the reliability perturbation function given by Eq. (48), it can be shown that the relative weights for e_1 and e_2 when $\sigma \to 0$ is given by

$$\bar{w}_1 = w_1 b_2/(w_1 b_2 + w_2 b_1) \quad \text{and} \quad \bar{w}_2 = w_2 b_1/(w_1 b_2 + w_2 b_1) \tag{49}$$

Note from Eq. (49) that a special case is $\bar{w}_i = w_i / \sum_{l=1}^{L} w_l$ if all L pieces of evidence are deemed to be equally reliable, whether they are in high conflict or not.

The above discussions lead to the following more detailed definition of the *ER* rule to combine joint support from two pieces of independent evidence that are each fully reliable,

$$p_{\theta,e(2)} = \begin{cases} 0 & \theta = \emptyset \\ \sum_{B\cap C=\theta} p_{B,1} p_{C,2} / (1 - \sum_{B\cap C=\emptyset} p_{B,1} p_{C,2}) & \theta \subseteq \Theta, \sigma > \tau \\ \hat{m}_{\theta,e(2)}(\sigma) / \sum_{D\subseteq\Theta} \hat{m}_{D,e(2)}(\sigma) & \theta \subseteq \Theta, \tau \geqslant \sigma > 0 \\ \bar{w}_1 p_{\theta,1} + \bar{w}_2 p_{\theta,2} & \theta \subseteq \Theta, \sigma = 0 \end{cases} \tag{50}$$

$$\hat{m}_{\theta,e(2)}(\sigma) = \left[\left(1 - r_2(\sigma)\right) m_{\theta,1} + \left(1 - r_1(\sigma)\right) m_{\theta,2}\right] + \sum_{B\cap C=\theta} m_{B,1} m_{C,2} \quad \forall \theta \subseteq \Theta \tag{51}$$

In Eq. (51) $r_i(\sigma)$ is calculated by Eq. (48). In Eq. (50), the second formula is the application of Corollary 6, the third formula is the application of the *ER* rule given by Corollary 3 for a specific reliability perturbation function $r_i(\sigma)$, and formula (4) is based on the above conjecture of the weighted average rule. While the conjecture needs to be proved rigorously, we will conduct simulation study in next section to show that formula (3) converges to formula (2) when $\sigma \to \tau$ and to formula (4) when $\sigma \to 0$. As such, the special case of the *ER* rule given by Eqs. (50) and (51) completes and enhances Dempster's rule.

Resolution of high conflict among evidence depends on how evidence is gathered and how the *DM* responds to a situation where evidence is in high conflict. It is important to examine these key factors in order to decide what reliability perturbation function should be used to analyse and resolve conflict. As illustrated by the simulation study reported in next section, for example, in the function shown in Eq. (48), the choice of the parameters b_i, τ and k requires specific knowledge in a problem domain and depends on the interpretation of high conflict among evidence. In other words, there is no unique or optimal strategy for resolution of high conflict among evidence.

5.4. Simulation study for conflict resolution

The proposed reliability perturbation analysis can be used to conduct simulation study to demonstrate how the *ER* rule can be used to facilitate resolution of high conflict among evidence. Since Zadeh's "counter-intuitive" example [45–47] is so influential that it has put Dempster's rule in doubt over the last several decades, it seems necessary to analyse this example in detail to clarify the doubt.

Table 6 show a generalised version of Zadeh's example with $1 \geqslant \delta \geqslant 0$ in the table. Zadeh's example is a special case with $\delta = 0.01$. Note from Eq. (47) that there is $\sigma = \delta^2$ for the example of Table 6. In the long debate over the validity of Dempster's rule, there have been two main different types of view, one opposing it and the other supporting it. The main point of the opposing view is that how on earth proposition B could be surely supported when each of the two pieces of evidence only points to the proposition with such a low probability of 0.01. In this view, the probability of 0.01 may already mean "unlikely". On the other hand, 0.01 may still mean "likely" in the supporting view. Is either of these views correct and why? In the *ER* rule given by Eqs. (50) and (51), the answer is clear. If $\delta = 0.01$ means "unlikely", then $\sigma = \delta^2 = 10^{-4}$ must mean "highly unlikely"; so it is inappropriate to apply Dempster's rule in this case. On the other hand, if $\delta = 0.01$ means "likely", does a probability of 10^{-4} still mean "likely"? If not, it is inappropriate to apply Dempster's rule either.

The resolution of the above conflict requires extra information from the *DM*. Without loss of generality, suppose $\tau = 0.01$, $k = 1$, $b_1 = b_2 = 1$ and $w_1 = w_2 = 1$. For the example shown in Table 6, these parameter settings mean that: (1) e_1 and e_2 are deemed to be equally reliable although in conflict (or $b_1 = b_2 = 1$); (2) their reliability becomes doubtful at $\delta = 0.1$ (or $\sigma = \delta^2 = 0.01 = \tau$) or less; (3) the most severe doubt is deemed to happen at $\delta = 0.01$ (or $\sigma = \delta^2 = 0.0001 = \tau^2$) when both e_1 and e_2 become unlikely to support proposition B; (4) any loss of reliability for either e_1 or e_2 is no more than 1% (or $b_1\tau^k = b_2\tau^k = 0.01$). For these typical settings, the changes of $p_{A,e(2)}$, $p_{B,e(2)}$ and $p_{C,e(2)}$ with respect to δ are shown in Fig. 1, generated by using the third formula in Eq. (50) for $\tau \geqslant \sigma > 0$.

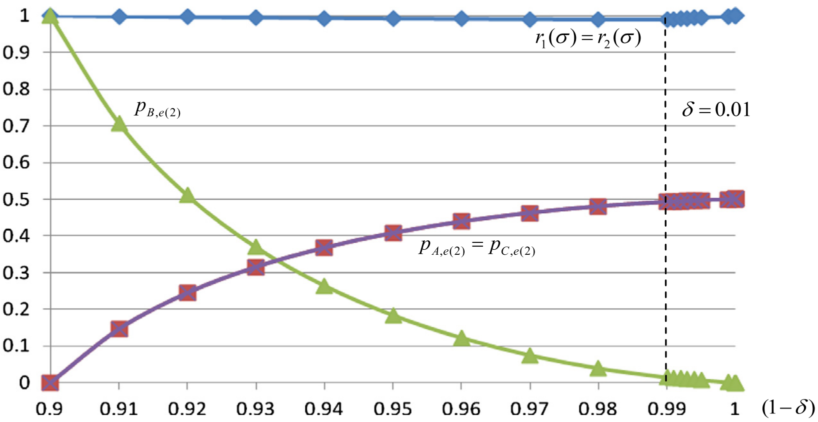

Fig. 1. Convergence of beliefs with δ for $b_1=b_2=1$.

As shown but hardly visible in Fig. 1, because b_1 and b_2 are both set to be 1, $r_1(\sigma)$ and $r_2(\sigma)$ are only perturbed slightly with $1 \geqslant r_1(\sigma)=r_2(\sigma) \geqslant 0.99$ ($\sigma=\delta^2$), decreasing when $1-\delta \in [0.9, 0.99]$ ($\delta \in [0.01, 0.1]$) monotonically, and increasing for $1-\delta \in [0.99, 1]$ ($\delta \in [0, 0.01]$) monotonically. Note that at $\delta=0.01$ (or $\sigma=0.0001$), $p_{B,e(2)}=0.0149$ and $p_{A,e(2)}=p_{C,e(2)}=0.4925$, and when $\delta \to 0$ they converge to the same result as generated by using the weighted average operation (the fourth formula in Eq. (50) for $\sigma=0$), or $p_{B,e(2)}=0$ and $p_{A,e(2)}=p_{C,e(2)}=0.5$, with $\bar{w}_1=\bar{w}_2=0.5$ in this case.

Various types of simulation study can be conducted to validate the *ER* rule by changing the settings of the parameters. For example, Fig. 2.1 shows the perturbation of the reliability $r_1(\sigma)(r_2(\sigma))$ of $e_1(e_2)$ with respect to both $b_1(b_2=b_1)$ and δ with $\tau=0.01$, $k=1$, $w_1=w_2=1$, and $r_1(\sigma)=r_2(\sigma)$. The change of the reliability is negligible for $b_1 \leqslant 1$, becomes greater for large b_1, and is the greatest for $b_1=100=1/\tau$.

Figs. 2.2 and 2.3 show respectively the changes of $p_{A,e(2)}(p_{C,e(2)})$ and $p_{B,e(2)}$ corresponding to Fig. 2.1. It is worth noting that the curves in Figs. 2.2–2.3 can be explained as different attitudes towards individual support for A or C and collective support for B when the two pieces of evidence are in high conflict. A large b_1 (e.g. $100 \geqslant b_1 > 0.3$) can mean an optimistic attitude to individual support and a pessimistic attitude to collective support; a small b_1 (e.g. $0.3 > b_1 \geqslant 0$) can mean exactly the opposite; a medium b_1 of around 0.3 can mean a neutral attitude to both in this example. The curves in both Figs. 2.2 and 2.3 for $b_1=0.01$ and 100 envelope all the combined beliefs for A, B and C for all the possible values of b_1. For example, at $\delta=0.01$, $p_{A,e(2)}(p_{C,e(2)})$ is between 0.33 for $b_1=0.01$ and 0.495 for $b_1=100$, and $p_{B,e(2)}$ is between 0.01 for $b_1=100$ and 0.34 for $b_1=0.01$, with $p_{A,e(2)}+p_{B,e(2)}+p_{C,e(2)}=1$.

The results of Figs. 2.1–2.3 can be used to explain how the conflict in Zadeh's example, where $\delta=0.01$, could be resolved in different ways. For example, the use of the *PCR5* rule leads to $p_{B,e(2)}=0.026$ and $p_{A,e(2)}=p_{C,e(2)}=0.487$ [14], which is about the same as the results of $p_{B,e(2)}=0.0262$ and $p_{A,e(2)}=p_{C,e(2)}=0.4869$ generated by using the *ER* rule with $b_1=b_2=0.3$. This is not surprising as the *PCR5* rule employs a proportional strategy to allocate beliefs in conflict, which is similar to a neutral attitude towards individual and collective support, although the *PCR5* rule is not a probabilistic reasoning process.

Fig. 3 shows the changes of the combined beliefs with respect to δ ($\sigma=\delta^2$) for the different weights of $w_1=1$ and $w_2=0.5$ with $\tau=0.01$, $k=1$ and $b_1=b_2=1$. In this case, e_1 is given a relative weight that is twice as large as that given to e_2, so from Eq. (49) we get $\bar{w}_1=0.6667$ and $\bar{w}_2=0.3333$. As shown in Fig. 3, when $\delta \to 0$ ($\sigma \to 0$), $p_{A,e(2)}$, $p_{B,e(2)}$ and $p_{C,e(2)}$ converge to 0.6667, 0 and 0.3333 respectively, just as expected by the weighted average operation of the fourth formula in Eq. (50).

Fig. 4 show the same case as Fig. 3 except that $b_1=1$ and $b_2=2$, so that e_2 is put in doubt twice as much as e_1, leading to $\bar{w}_1=0.8$ and $\bar{w}_2=0.2$ from Eq. (49). Indeed we have $p_{A,e(2)}=0.8$, $p_{B,e(2)}=0$ and $p_{C,e(2)}=0.2$ when $\delta \to 0$ ($\sigma \to 0$).

Finally, it is worth illustrating how the *ER* rule behaves for conflict resolution in Zadeh's example at $\delta=0.01$ with various speeds of convergence (or various k values). The results are shown in Table 7, where it is assumed that $b_1=b_2=1$, $\tau=0.01$ and $w_1=w_2=1$. As shown in Table 7, for various k, $r_1(\sigma)$ $(=r_2(\sigma))$ is perturbed very differently, but the values of $p_{A,e(2)}$, $p_{B,e(2)}$ and $p_{C,e(2)}$ are kept almost unchanged, showing that the convergence is fairly robust. It can be shown that this is also the case for $b_1, b_2 \geqslant 1$. For small b_1 (b_2), the values of $p_{A,e(2)}$, $p_{B,e(2)}$ and $p_{C,e(2)}$ are more sensitive to the value of k, as shown in Fig. 5 for $0.1 \geqslant b_1=b_2 \geqslant 0.01$. Note that $p_{C,e(2)}=p_{A,e(2)}$ and $p_{B,e(2)}=1-p_{A,e(2)}-p_{C,e(2)}$. This shows that it is important to choose an appropriate reliability perturbation function for conflict resolution if e_1 and e_2 are still deemed to be highly reliable when in high conflict.

In the above analyses, it should be noted that for small b_1 (or b_2) the change of reliability for e_1 (or e_2) is negligible, so the reliability is just perturbed rather than necessarily reduced. In this regard, the use of the *ER* rule as shown in Eq. (50) does not mean that reliability for any evidence must be sacrificed for conflict resolution. Although the reliability perturbation analysis leads to the establishment of the *ER* rule for combining pieces of fully reliable evidence that are highly

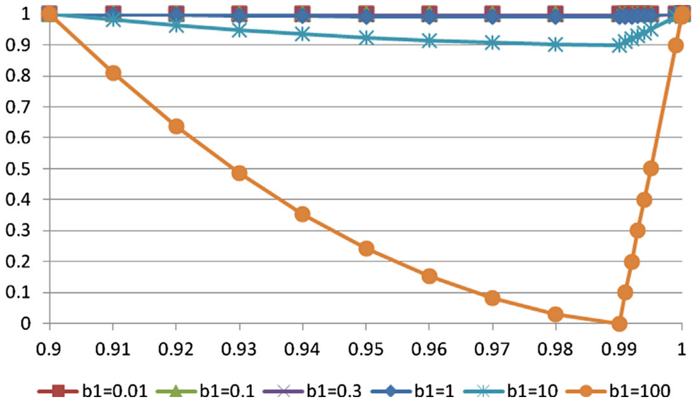

Fig. 2.1. Perturbation of reliability $r_1(\sigma)$ $(r_2(\sigma) = r_1(\sigma))$ with δ for various b_1 (b_2).

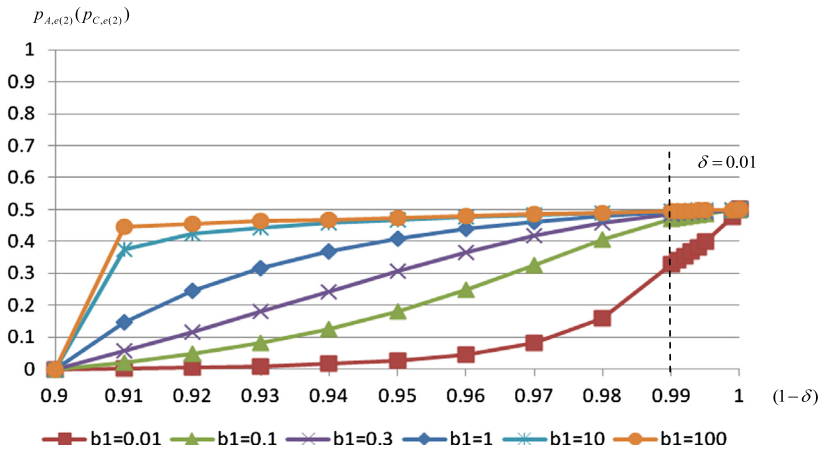

Fig. 2.2. Convergence of $p_{A,e(2)}$ $(p_{C,e(2)})$ with δ for various b_1 $(= b_2)$.

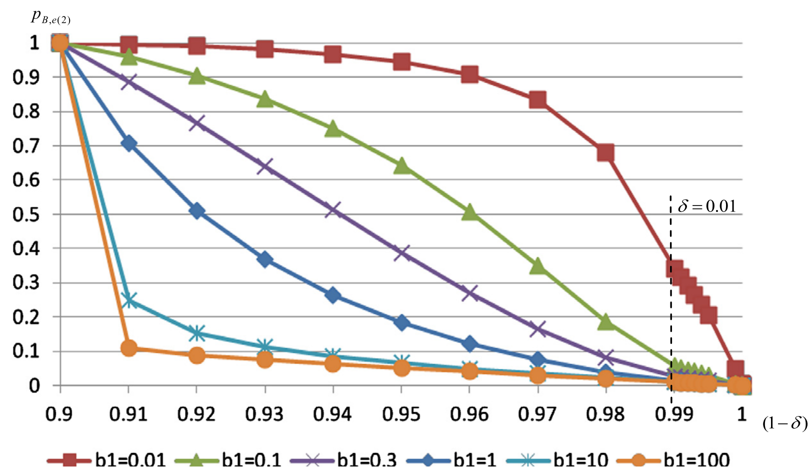

Fig. 2.3. Convergence of $p_{B,e(2)}$ with respect to δ for various b_1 $(= b_2)$.

or completely conflicting, the analysis is not a rigorous proof that the *ER* rule is the only legitimate probabilistic reasoning process for combining this type of evidence. To find such a proof or identify any other probabilistic reasoning processes is a topic for further research.

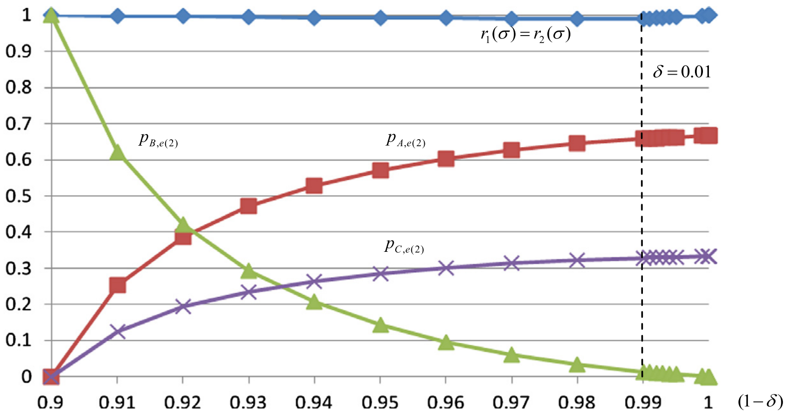

Fig. 3. Convergence of beliefs with δ for $b_1 = b_2 = 1$, $w_1 = 1$ and $w_2 = 0.5$.

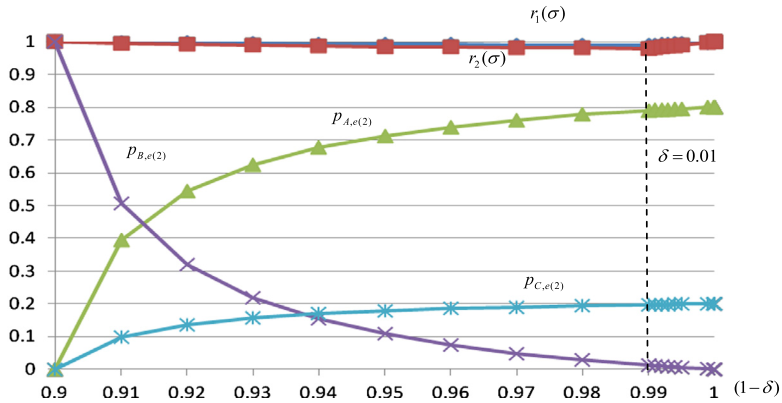

Fig. 4. Convergence of beliefs with δ for $b_1 = 1$, $b_2 = 2$, $w_1 = 1$ and $w_2 = 0.5$.

Table 7
Convergence of beliefs for various k when $b_i \geqslant 1$ ($b_i = 1$ is used in this table).

k	1	0.8	0.6	0.4	0.2	0.1
$r_1(\sigma)$ $(= r_2(\sigma))$	0.9900	0.9749	0.9369	0.8415	0.6019	0.369
$p_{A,e(2)}$	0.4925	0.4940	0.4946	0.4948	0.4949	0.495
$p_{B,e(2)}$	0.0149	0.0120	0.0108	0.0103	0.0102	0.010
$p_{C,e(2)}$	0.4925	0.4940	0.4946	0.4948	0.4949	0.495

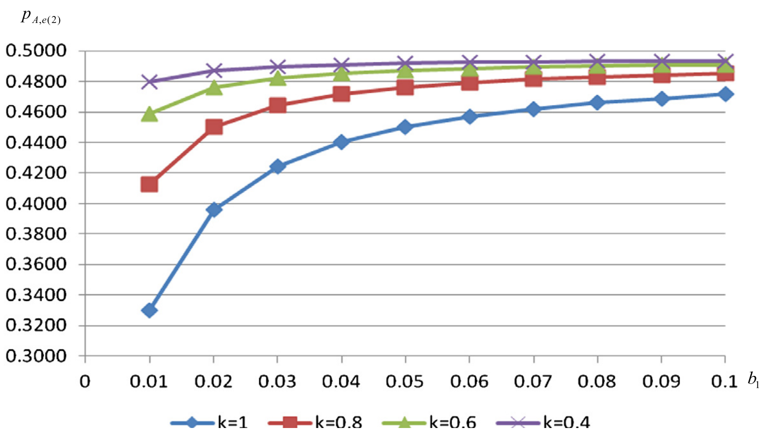

Fig. 5. Convergence of beliefs at $\delta = 0.01$ for various b_1 and k.

6. Discussions and conclusion

In this paper, the unique *Evidential Reasoning* (*ER*) rule was established, its inherent properties were explored in detail, and comprehensive numerical and simulation studies were conducted to facilitate its comparison with other existing evidence combination rules. The *ER* rule reveals that the combined degree of joint support for a proposition from two pieces of independent evidence constitutes two parts in general: the bounded sum of their individual support and the orthogonal sum of their collective support. The *ER* rule is based on the orthogonal sum operation and thus inherits the basic properties of being associative and commutative, so that it can be used to combine multiple pieces of evidence in any order without changing the final results. It was proven that the *ER* rule satisfies four synthesis axioms that any rational probabilistic reasoning process should follow. It was shown that the *ER* algorithm is a special case of the *ER* rule when the reliability of evidence is equal to its weight that is normalised to be relative to each other among all pieces of evidence. It was proven that Dempster's rule is a special case of the *ER* rule as well when each piece of evidence is fully reliable. It was also shown that the *ER* rule completes and enhances Dempster's rule by identifying how to combine multiple pieces of fully reliable evidence that are highly or completely conflicting through the reliability perturbation analysis proposed in this paper.

The *ER* rule advances *D–S* theory of evidence and the *ER* approach. On one hand, the *Belief Distribution* (*BD*) introduced in *D–S* theory was employed as a basis to profile a piece of evidence; on the other hand, the basic probability assignment method proposed in the *ER* approach was generalised to construct a novel *Weighted Belief Distribution* (*WBD*) and a *WBD* with *Reliability* (*WBDR*) for characterising evidence equivalently in complement of *BD*. In the process of constructing *WBD* (or *WBDR*), the specificity of evidence is kept intact. It is the implementation of the orthogonal sum operation on *WBD*s (or *WBDR*s) that leads to the establishment of the *ER* rule. As such, the *ER* rule constitutes a generic conjunctive probabilistic reasoning process, or a generalised Bayesian inference process, which is applicable to combine multiple pieces of independent evidence with different weights and reliabilities in a wide range of areas such as multiple criteria decision analysis [32,34,35] and information fusion [14]. The proposed reliability perturbation analysis is based on both the interpretation of conflict that if conflict between two information sources is high at least one of them may be unreliable, and the conjecture that if two pieces of evidence are in complete conflict the weighted average rule should be applied to combine their *BD*s.

The above conjecture is based on the observation that the conjunction of two pieces of evidence will be empty if they are completely conflicting, which is a special case of mutual exclusiveness. While it is widely accepted that the combination of two pieces of mutually exclusive evidence should be additive, the conjecture needs to be proven rigorously, which is beyond the scope of this paper and requires further research. Nevertheless, the interpretation of conflict is well researched in literature, and the reliability perturbation analysis led to the completion and enhancement of Dempster's rule by exploring doubt over reliability rather than assuming that the reliability of any evidence has to be sacrificed for conflict resolution. The development of actual strategies for conflict resolution depends on how evidence is gathered and whether and by how much the decision maker may have doubt over the reliability of the evidence in high conflict. Finally, it should be noted that the *ER* rule is applicable only to conjunctive combination of independent evidence. Other rules need to be established to combine dependent evidence, or for non-conjunctive or mixed combination of multiple pieces of evidence that may or may not be independent [7].

Acknowledgements

This research is partly supported by European Commission under the grant No. EC-GPF-314836. The constructive comments provided by three anonymous referees and the associate editor have helped improve the quality of this paper to its current standard.

Appendix A

A.1. Proof of Lemma 1

Proof. From Eqs. (16a), (16b), (19c) and (19d) and as $m_{P(\Theta),1} = 1 - w_1$ and $m_{P(\Theta),2} = 1 - w_2$, it is clear by rearranging the terms in Eqs. (16a) and (16b) that there are $m_{n,e(2)} = k\hat{m}_{n,e(2)}$ and $m_{\Theta,e(2)} = k\hat{m}_{\Theta,e(2)}$. As $\sum_{t=1}^{N} m_{t,e(2)} + m_{\Theta,e(2)} + m_{P(\Theta),e(2)} = 1$ [41], we get $1 - m_{P(\Theta),e(2)} = k(\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)})$. From Eq. (17a), we then have

$$p_{n,e(2)} = \frac{m_{n,e(2)}}{1 - m_{P(\Theta),e(2)}} = \frac{k\hat{m}_{n,e(2)}}{k(\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)})} = \frac{\hat{m}_{n,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}} = k_1\hat{m}_{n,e(2)} \tag{A.1.1}$$

Similarly, from Eq. (17b), we get

$$p_{\Theta,e(2)} = \frac{m_{\Theta,e(2)}}{1 - m_{P(\Theta),e(2)}} = \frac{k\hat{m}_{\Theta,e(2)}}{k(\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)})} = \frac{\hat{m}_{\Theta,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}} = k_1\hat{m}_{\Theta,e(2)} \quad \square \tag{A.1.2}$$

A.2. Proof of Theorem 1

Proof. Let $m_{n,i} = w_i p_{n,i} (n = 1, \ldots, N)$ and $m_{\Theta,i} = w_i p_{\Theta,i}$ as the probability masses or the degrees of support for θ_n and Θ from the ith piece of evidence. Construct the *WBD* of the ith piece of evidence as m_i defined by Eq. (28) for all $i = 1, \ldots, L$.

Note that $\theta_n \cap \Theta = \theta_n$ and $\theta_n \cap P(\Theta) = \theta_n$ for any $n = 1, \ldots, N$, and $\Theta \cap P(\Theta) = \Theta$. For $i = 2$, first apply the orthogonal sum operation to combine m_1 and m_2, leading to the combined probability masses or degrees of support for θ_n and Θ, denoted by $\hat{m}_{n,e(2)}$ and $\hat{m}_{\Theta,e(2)}$, and also resulting in the combined residual support for $P(\Theta)$, denoted by $\hat{m}_{P(\Theta),e(2)}$

$$\begin{aligned}\hat{m}_{n,e(2)} &= m_{n,1}m_{n,2} + m_{n,1}m_{\Theta,2} + m_{\Theta,1}m_{n,2} + m_{n,1}m_{P(\Theta),2} + m_{P(\Theta),1}m_{n,2}\\ &= \big[(1-w_2)m_{n,1} + (1-w_1)m_{n,2}\big] + [m_{n,1}m_{n,2} + m_{n,1}m_{\Theta,2} + m_{\Theta,1}m_{n,2}], \quad n = 1, \ldots, N\end{aligned} \tag{A.2.1}$$

$$\begin{aligned}\hat{m}_{\Theta,e(2)} &= m_{\Theta,1}m_{\Theta,2} + m_{\Theta,1}m_{P(\Theta),2} + m_{P(\Theta),1}m_{\Theta,2}\\ &= \big[(1-w_2)m_{\Theta,1} + (1-w_1)m_{\Theta,2}\big] + [m_{\Theta,1}m_{\Theta,2}], \quad n = 1, \ldots, N\end{aligned} \tag{A.2.2}$$

$$\hat{m}_{P(\Theta),e(2)} = m_{P(\Theta),1}m_{P(\Theta),2} = (1-w_1)(1-w_2) \tag{A.2.3}$$

Now normalise $\hat{m}_{n,e(2)}$, $\hat{m}_{\Theta,e(2)}$ and $\hat{m}_{P(\Theta),e(2)}$ by a factor k_2, denoted by

$$m_{n,e(2)} = k_2\hat{m}_{n,e(2)}, \qquad m_{\Theta,e(2)} = k_2\hat{m}_{\Theta,e(2)} \quad \text{and} \quad m_{P(\Theta),e(2)} = k_2\hat{m}_{P(\Theta),e(2)} \tag{A.2.4}$$

so that all the combined probability masses are added to one, or

$$\sum_{t=1}^{N} m_{t,e(2)} + m_{\Theta,e(2)} + m_{P(\Theta),e(2)} = 1 \tag{A.2.5}$$

leading to

$$k_2 = \left(\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)} + \hat{m}_{P(\Theta),e(2)}\right)^{-1} \tag{A.2.6}$$

The result generated by combining two pieces of evidence represented by m_1 and m_2 can therefore be given by the following combined *WBD*:

$$m_{e(2)} = \big\{(\theta_n, m_{n,e(2)}),\ n = 1, \ldots, N;\ (\Theta, m_{\Theta,e(2)});\ \big(P(\Theta), m_{P(\Theta),e(2)}\big)\big\} \tag{A.2.7}$$

If there is another piece of evidence, such as m_3, then $m_{e(2)}$ can be combined with m_3 to generate a new combined *WBD* in the same fashion as shown by Eqs. (A.2.1) to (A.2.6) with $m_{e(2)}$ treated as m_1 and m_3 as m_2. This process can be repeated until all *WBD*s are combined.

To terminate the process, without loss of generality, suppose there are two pieces of evidence for combination in total, or $w_1 + w_2 = 1$ for example. From Eqs. (27), (A.2.4) and (A.2.5), the combined *BD* is then given as follows:

$$e(2) = \big\{(\theta_n, p_{n,e(2)}),\ n = 1, \ldots, N;\ (\Theta, p_{\Theta,e(2)})\big\} \tag{A.2.8}$$

$$p_{n,e(2)} = \frac{m_{n,e(2)}}{1 - m_{P(\Theta),e(2)}} = \frac{m_{n,e(2)}}{\sum_{t=1}^{N} m_{t,e(2)} + m_{\Theta,e(2)}} = \frac{\hat{m}_{n,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}}, \quad n = 1, \ldots, N \tag{A.2.9}$$

$$p_{\Theta,e(2)} = \frac{m_{\Theta,e(2)}}{1 - m_{P(\Theta),e(2)}} = \frac{m_{\Theta,e(2)}}{\sum_{t=1}^{N} m_{t,e(2)} + m_{\Theta,e(2)}} = \frac{\hat{m}_{\Theta,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}} \tag{A.2.10}$$

$$\begin{aligned}\sum_{n=1}^{N} p_{n,e(2)} + p_{\Theta,e(2)} &= \sum_{n=1}^{N}\left\{\frac{\hat{m}_{n,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}}\right\} + \frac{\hat{m}_{\Theta,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}}\\ &= \frac{\sum_{n=1}^{N} \hat{m}_{n,e(2)} + \hat{m}_{\Theta,e(2)}}{\sum_{t=1}^{N} \hat{m}_{t,e(2)} + \hat{m}_{\Theta,e(2)}} = 1\end{aligned} \tag{A.2.11}$$

From Lemma 1 and Eqs. (A.2.9) and (A.2.10), we conclude that the algorithm composed of Eqs. (A.2.1)–(A.2.10) is the same as the original *ER* algorithm given by Eqs. (12)–(18). □

A.3. Proof of Theorem 2

Proof. First, we need to prove that Eqs. (25), (29) and (30) represent the combined *WBD* generated by applying the orthogonal sum to combine the following two *WBD*s:

$$m_i = \{(\theta, m_{\theta,i}),\ \forall\theta \subseteq \Theta;\ (P(\Theta), m_{P(\Theta),i})\} \quad \text{for } i = 1, 2 \tag{A.3.1}$$

with $m_{\theta,i} = w_i p_{\theta,i}$ $\forall\theta \subseteq \Theta$ and $m_{P(\Theta),i} = 1 - w_i$. The orthogonal sum operation of m_1 and m_2 for any $\theta \subseteq \Theta$ and $P(\Theta)$ without normalisation leads to

$$\begin{aligned}\hat{m}_{\theta,e(2)} &= \sum_{B\cap C=\theta;\, B,C\subseteq\Theta} m_{B,1}m_{C,2} + [m_{\theta,1}m_{P(\Theta),2} + m_{P(\Theta),1}m_{\theta,2}] \\ &= [(1-w_2)m_{\theta,1} + (1-w_1)m_{\theta,2}] + \sum_{B\cap C=\theta;\, B,C\subseteq\Theta} m_{B,1}m_{C,2}, \quad \forall\theta \subseteq \Theta\end{aligned} \tag{A.3.2}$$

$$\hat{m}_{P(\Theta),e(2)} = m_{P(\Theta),1}m_{P(\Theta),2} = (1-w_1)(1-w_2) \tag{A.3.3}$$

Since $m_{\emptyset,e(2)} = 0$, let $m_{\theta,e(2)} = k\hat{m}_{\theta,e(2)}$ $\forall\theta \subseteq \Theta$ and $m_{P(\Theta),e(2)} = k\hat{m}_{P(\Theta),e(2)}$. We can now normalise the above combined probability masses so that they are still added to one, or

$$\sum_{D\subseteq\Theta} k\hat{m}_{D,e(2)} + k\hat{m}_{P(\Theta),e(2)} = 1, \quad \text{or} \quad k = \frac{1}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}} \tag{A.3.4}$$

So

$$m_{\theta,e(2)} = \frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}} \quad \forall\theta \subseteq \Theta, \quad \text{and} \quad m_{P(\Theta),e(2)} = \frac{\hat{m}_{P(\Theta),e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}} \tag{A.3.5}$$

Since each term is non-negative in Eqs. (A.3.2) and (A.3.3), there are $\hat{m}_{\theta,e(2)} \geqslant 0$ $\forall\theta \subseteq \Theta$ and $\hat{m}_{P(\Theta),e(2)} \geqslant 0$. From Eq. (A.3.5), we then get

$$\begin{aligned}&m_{\theta,e(2)} \geqslant 0 \quad \forall\theta \subseteq \Theta, \qquad m_{P(\Theta),e(2)} \geqslant 0 \quad \text{and} \\ &\sum_{\theta\subseteq\Theta} m_{\theta,e(2)} + m_{P(\Theta),e(2)} = \sum_{\theta\subseteq\Theta}\left\{\frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}}\right\} + \frac{\hat{m}_{P(\Theta),e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}} \\ &\qquad = \frac{\sum_{\theta\subseteq\Theta}\hat{m}_{\theta,e(2)} + \hat{m}_{P(\Theta),e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}} = 1\end{aligned} \tag{A.3.6}$$

which in turn means that there must be

$$m_{\theta,e(2)} \leqslant 1 \quad \forall\theta \subseteq \Theta \quad \text{and} \quad m_{P(\Theta),e(2)} \leqslant 1 \tag{A.3.7}$$

Note that the normalisation in Eq. (A.3.5) is equivalent to that used in Eq. (5) since the orthogonal sum of the two *WBD*s always leads to the following equation

$$\left(\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)} + \hat{m}_{P(\Theta),e(2)}\right) + \sum_{B\cap C=\emptyset} m_{B,1}m_{C,2} = \sum_{B\cap C\neq\emptyset} m_{B,1}m_{C,2} + \sum_{B\cap C=\emptyset} m_{B,1}m_{C,2} = 1 \tag{A.3.8}$$

Finally, from Eqs. (27), (A.3.2), (A.3.3) and (A.3.4), the combined degrees of belief generated from the two *BD*s e_1 and e_2 are given by

$$p_{\theta,e(2)} = \frac{m_{\theta,e(2)}}{1 - m_{P(\Theta),e(2)}} = \frac{k\hat{m}_{\theta,e(2)}}{1 - k\hat{m}_{P(\Theta),e(2)}} = \frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)}}, \quad \forall\theta \subseteq \Theta \tag{A.3.9}$$

Since $\hat{m}_{\theta,e(2)} \geqslant 0$ $\forall\theta \subseteq \Theta$, from Eq. (A.3.9) we have $p_{\theta,e(2)} \geqslant 0$ $\forall\theta \subseteq \Theta$ and there is

$$\sum_{\theta\subseteq\Theta} p_{\theta,e(2)} = \sum_{\theta\subseteq\Theta}\left\{\frac{\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)}}\right\} = \frac{\sum_{\theta\subseteq\Theta}\hat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta}\hat{m}_{D,e(2)}} = 1 \tag{A.3.10}$$

which in turn means that there must be $p_{\theta,e(2)} \leqslant 1$ $\forall\theta \subseteq \Theta$. □

A.4. Proof of Corollary 1

Proof. For $i=2$, since $m_{\theta,e(1)}=m_{\theta,1}$, $m_{P(\Theta),e(1)}=m_{P(\Theta),1}$, Eq. (32) is equal to Eq. (30). Therefore, Eq. (31) is true with $0\leqslant m_{\theta,e(2)}\leqslant 1$ for any $\theta\subseteq\Theta$, $0\leqslant m_{P(\Theta),e(2)}\leqslant 1$ and $\sum_{\theta\subseteq\Theta} m_{\theta,e(2)}+m_{P(\Theta),e(2)}=1$ according to Theorem 2.

Suppose for $i=l$, Eqs. (31), (32) and (33) are also true, that is

$$m_{\theta,e(l)}=[m_1\oplus\cdots\oplus m_l](\theta) \tag{A.4.1}$$

with $0\leqslant m_{\theta,e(l)}\leqslant 1$ for any $\theta\subseteq\Theta$, $0\leqslant m_{P(\Theta),e(l)}\leqslant 1$ and $\sum_{\theta\subseteq\Theta} m_{\theta,e(l)}+m_{P(\Theta),e(l)}=1$.

For $i=l+1$, since the orthogonal sum of *WBDs* is independent of the order in which they are combined, we have

$$m_{\theta,e(l+1)}=[m_1\oplus\cdots\oplus m_{l+1}](\theta)=\big[[m_1\oplus\cdots\oplus m_l]\oplus m_{l+1}\big](\theta) \tag{A.4.2}$$

The above equation means that combining $(l+1)$ *WBDs* is equal to combining the first l *WBDs* with the $(l+1)$th *WBD*. The orthogonal sum of $m_{e(l)}$ and m_{l+1} without normalisation leads to

$$\begin{aligned}\hat{m}_{\theta,e(l+1)}&=\sum_{B\cap C=\theta;B,C\subseteq\Theta} m_{B,e(l)}m_{C,l+1}+[m_{\theta,e(l)}m_{P(\Theta),l+1}+m_{P(\Theta),e(l)}m_{\theta,l+1}]\\&=\big[(1-w_{l+1})m_{\theta,e(l)}+m_{P(\Theta),e(l)}m_{\theta,l+1}\big]+\sum_{B\cap C=\theta;B,C\subseteq\Theta} m_{B,e(l)}m_{C,l+1},\quad\forall\theta\subseteq\Theta\end{aligned} \tag{A.4.3}$$

$$\hat{m}_{P(\Theta),e(l+1)}=m_{P(\Theta),e(l)}m_{P(\Theta),l+1}=(1-w_{l+1})m_{P(\Theta),e(l)} \tag{A.4.4}$$

Since $m_{\emptyset,e(l+1)}=0$, let $m_{\theta,e(l+1)}=k\hat{m}_{\theta,e(l+1)}$ $\forall\theta\subseteq\Theta$ and $m_{P(\Theta),e(l+1)}=k\hat{m}_{P(\Theta),e(l+1)}$. Normalise the combined probability masses given by Eqs. (A.4.3) and (A.4.4) so that they are still added to one, or

$$\sum_{B\subseteq\Theta} k\hat{m}_{B,e(l+1)}+k\hat{m}_{P(\Theta),e(l+1)}=1,\quad\text{so}\quad k=\frac{1}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}} \tag{A.4.5}$$

We finally have

$$\begin{aligned}m_{\theta,e(l+1)}&=\frac{\hat{m}_{\theta,e(l+1)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}\quad\forall\theta\subseteq\Theta\quad\text{and}\\m_{P(\Theta),e(l+1)}&=\frac{\hat{m}_{P(\Theta),e(l+1)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}\end{aligned} \tag{A.4.6}$$

Since every term in Eqs. (A.4.3) and (A.4.4) is non-negative, there must be $\hat{m}_{\theta,e(l+1)}\geqslant 0$ $\forall\theta\subseteq\Theta$ and $\hat{m}_{P(\Theta),e(l+1)}\geqslant 0$. From Eq. (A.4.6), we have $m_{\theta,e(l+1)}\geqslant 0$ $\forall\theta\subseteq\Theta$, $m_{P(\Theta),e(l+1)}\geqslant 0$ and

$$\begin{aligned}\sum_{\theta\subseteq\Theta} m_{\theta,e(l+1)}+m_{P(\Theta),e(l+1)}&=\sum_{\theta\subseteq\Theta}\left\{\frac{\hat{m}_{\theta,e(l+1)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}\right\}+\frac{\hat{m}_{P(\Theta),e(l+1)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}\\&=\frac{\sum_{\theta\subseteq\Theta}\hat{m}_{\theta,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(l+1)}+\hat{m}_{P(\Theta),e(l+1)}}=1\end{aligned} \tag{A.4.7}$$

which in turn means that $m_{\theta,e(l+1)}\leqslant 1$ $\forall\theta\subseteq\Theta$ and $m_{P(\Theta),e(l+1)}\leqslant 1$. □

A.5. Proof of Corollary 2

Proof. Corollary 1 shows that the results generated from the orthogonal sum of the L *WBDs* m_i are given by Eqs. (31), (32) and (33) for $i=L$. From Eqs. (27), (A.4.3) and (A.4.5), the combined degrees of belief generated from e_i $(i=1,\dots,L)$ are given by

$$p_{\theta,e(L)}=\frac{m_{\theta,e(L)}}{1-m_{P(\Theta),e(L)}}=\frac{k\hat{m}_{\theta,e(L)}}{1-k\hat{m}_{P(\Theta),e(L)}}=\frac{\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(L)}},\quad\forall\theta\subseteq\Theta \tag{A.5.1}$$

with $\hat{m}_{\theta,e(L)}$ given by Eq. (A.4.3) for $l+1=L$, which is the same as Eq. (32) for $i=L$. From Corollary 1, we have $0\leqslant m_{\theta,e(i)}\leqslant 1$ $\forall\theta\subseteq\Theta$ for any $i=2,\dots,L$, so there is $\hat{m}_{\theta,e(L)}\geqslant 0$ $\forall\theta\subseteq\Theta$. From Eq. (A.5.1), we then have $p_{\theta,e(L)}\geqslant 0$ $\forall\theta\subseteq\Theta$ and

$$\sum_{\theta\subseteq\Theta} p_{\theta,e(L)}=\sum_{\theta\subseteq\Theta}\left\{\frac{\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(L)}}\right\}=\frac{\sum_{\theta\subseteq\Theta}\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(L)}}=1 \tag{A.5.2}$$

which in turn means that $p_{\theta,e(L)}\leqslant 1$ $\forall\theta\subseteq\Theta$. □

A.6. Proof of Theorem 3

Proof. Theorem 3 states that there will be no support for a proposition θ if no evidence supports θ or any proposition B that includes θ (or $B \cap \theta = \theta$).

For $L = 2$, suppose $p_{B,i} = 0$ for $i = 1, 2$ and any $B \subseteq \Theta$ with $B \cap \theta = \theta$. Note that if $B \cap C = \theta$ there must be $B \cap \theta = \theta$ and $C \cap \theta = \theta$. Therefore, $m_{\theta,1} = w_1 p_{\theta,1} = 0$, $m_{\theta,2} = w_2 p_{\theta,2} = 0$, and also $m_{B,1} = w_1 p_{B,1} = 0$ and $m_{C,2} = w_2 p_{C,2} = 0$ for any other $B, C \subseteq \Theta$ with $B \cap C = \theta$. From Eq. (30) we then have

$$\begin{aligned}\hat{m}_{\theta,e(2)} &= \left[(1 - w_2)m_{\theta,1} + (1 - w_1)m_{\theta,2}\right] + \sum_{B\cap C=\theta} m_{B,1}m_{C,2} \\ &= \left[(1 - w_2) \times 0 + (1 - w_1) \times 0\right] + \sum_{B\cap C=\theta} 0 \times 0 = 0\end{aligned}$$

From Eq. (29), there is $p_\theta = p_{\theta,e(2)} = \hat{m}_{\theta,e(2)} / \sum_{D\subseteq\Theta} \hat{m}_{D,e(2)} = 0$.

For $L = l$, suppose it is true that $p_\theta = p_{\theta,e(l)} = 0$ if $p_{B,i} = 0$ for any $B \subseteq \Theta$ with $B \cap \theta = \theta$ and $i = 1, \dots, l$. From Eq. (27) we then have $m_{\theta,e(l)} = p_{\theta,e(l)}(1 - m_{P(\Theta),e(l)}) = 0$.

For $L = l + 1$, note that $m_{\theta,l+1} = p_{\theta,l+1}(1 - m_{P(\Theta),l+1}) = 0$ and $m_{C,l+1} = p_{C,l+1}(1 - m_{P(\Theta),l+1}) = 0$ for any $C \subseteq \Theta$ with $C \cap \theta = \theta$. From Eq. (32), we have

$$\begin{aligned}\hat{m}_{\theta,e(l+1)} &= \left[(1 - w_{l+1})m_{\theta,e(l)} + m_{P(\Theta),e(l)}m_{\theta,l+1}\right] + \sum_{B\cap C=\theta} m_{B,e(l)}m_{C,l+1} \\ &= \left[(1 - w_{l+1}) \times 0 + m_{P(\Theta),e(l)} \times 0\right] + \sum_{B\cap C=\theta} m_{B,e(l)} \times 0 = 0\end{aligned}$$

From Eq. (34), we have $p_\theta = p_{\theta,e(l+1)} = \hat{m}_{\theta,e(l+1)} / \sum_{D\subseteq\Theta} \hat{m}_{D,e(l+1)} = 0$. □

A.7. Proof of Theorem 4

Proof. The condition $p_{B,i} = 0$ for all $i = 1, \dots, L$ and any $B \subseteq \Theta$ with $B \neq \theta$ means that $p_{D,i} = 0$ for all $i = 1, \dots, L$ and any $D \subseteq \Theta$ with $D \cap B = B \neq \theta$. From Theorem 3, there will be $p_B = 0$ for any $B \neq \theta$.

We now first need to prove that $\hat{m}_{\theta,e(L)} > 0$ as generated from Eq. (30) or Eq. (32) for $i = L$. For $L = 2$, since $m_{\theta,1} = w_1 p_{\theta,1} = w_1$, $m_{\theta,2} = w_2 p_{\theta,2} = w_2$ and w_1 and w_2 are given by Eq. (45), there must be $(1 - w_2)m_{\theta,1} + (1 - w_1)m_{\theta,2} = (1 - w_2)w_1 + (1 - w_1)w_2 > 0$. From Eq. (30), we have $\hat{m}_{\theta,e(2)} > 0$.

For $i = l$, suppose $\hat{m}_{\theta,e(l)} > 0$, so $m_{\theta,e(l)} > 0$ from Eq. (31). For $i = l + 1$, since $0 < w_{l+1} < 1$ from Eq. (45), there is $(1 - w_{l+1})m_{\theta,e(l)} > 0$. From Eq. (32), we then have $\hat{m}_{\theta,e(l+1)} > 0$. Thus, there must be $\hat{m}_{\theta,e(L)} > 0$ for any $L \geqslant 2$.

Note from Eq. (34) that we have $\hat{m}_{B,e(L)} = 0$ for any $B \subseteq \Theta$ with $B \neq \theta$ because $p_B = 0$ for any $B \neq \theta$. From Eq. (34) again, we finally have

$$p_\theta = p_{\theta,e(L)} = \frac{\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta} \hat{m}_{B,e(L)}} = \frac{\hat{m}_{\theta,e(L)}}{\hat{m}_{\theta,e(L)} + \sum_{B\subseteq\Theta, B\neq\theta} \hat{m}_{B,e(L)}} = \frac{\hat{m}_{\theta,e(L)}}{\hat{m}_{\theta,e(L)} + 0} = 1 \qquad \square$$

A.8. Proof of Theorem 5

Proof. The definition of Θ^+ means that it includes any proposition θ with a positive degree of belief from any evidence plus any of its subsets B (or $B \cap \theta = B$). Therefore, if $B \notin \Theta^+$, there will be no belief in B or any proposition D that includes B (or $D \cap B = B$), that is $p_{D,i} = 0$ for any $D \subseteq \Theta$ with $D \cap B = B \notin \Theta^+$ and all $i = 1, \dots, L$. From Theorem 3, we then get $p_B = 0$ for any $B \subseteq \Theta$ with $B \notin \Theta^+$. Since $p_B = p_{B,e(L)}$, from Eq. (34) we have $\hat{m}_{B,e(L)} = 0$ for any $B \subseteq \Theta$ with $B \notin \Theta^+$, so $\sum_{B\notin\Theta^+} \hat{m}_{B,e(L)} = 0$.

We now show that $\hat{m}_{P(\Theta),e(i)} > 0$ for any $i = 2, \dots, L$ given that Eq. (45) holds. For $L = 2$, from Eqs. (25) and (45), we have $\hat{m}_{P(\Theta),e(2)} = m_{P(\Theta),1}m_{P(\Theta),2} = (1 - w_1)(1 - w_2) > 0$.

For $L = l$, suppose $\hat{m}_{P(\Theta),e(l)} > 0$. From Eq. (31) we have $m_{P(\Theta),e(l)} > 0$. From Eqs. (33) and (45), we then get $\hat{m}_{P(\Theta),e(l+1)} = m_{P(\Theta),e(l)}m_{P(\Theta),l+1} = m_{P(\Theta),e(l)}(1 - w_{l+1}) > 0$, so $m_{P(\Theta),e(l+1)} > 0$ from Eq. (31). Then $\hat{m}_{P(\Theta),e(i)} > 0$ and $m_{P(\Theta),e(i)} > 0$ for any $i = 2, \dots, L$.

Since there must be $p_{\theta,i} > 0$ for at least one $i \in \{1, \dots, L\}$ with $\theta \in \Theta^+$, without loss of generality we can get the ith piece of evidence to be last combined without changing the final results due to the commutative property of the orthogonal sum operation, so that $p_{\theta,L} > 0$ and $m_{\theta,L} = w_L p_{\theta,L} > 0$ for $\theta \in \Theta^+$, leading to $m_{P(\Theta),e(L-1)}m_{\theta,L} > 0$. Since every other term in Eq. (32) for $i = L$ is non-negative, there must be $\hat{m}_{\theta,e(L)} > 0$ for $\theta \in \Theta^+$. As such, we get $\sum_{B\in\Theta^+} \hat{m}_{B,e(L)} > 0$. From Eq. (34), we have

$$\sum_{\theta\in\Theta^+} p_\theta = \sum_{\theta\in\Theta^+} p_{\theta,e(L)} = \sum_{\theta\in\Theta^+}\left\{\frac{\hat{m}_{\theta,e(L)}}{\sum_{B\subseteq\Theta}\hat{m}_{B,e(L)}}\right\} = \frac{\sum_{\theta\in\Theta^+}\hat{m}_{\theta,e(L)}}{\sum_{B\in\Theta^+}\hat{m}_{B,e(L)} + \sum_{B\notin\Theta^+}\hat{m}_{B,e(L)}}$$
$$= \frac{\sum_{\theta\in\Theta^+}\hat{m}_{\theta,e(L)}}{\sum_{B\in\Theta^+}\hat{m}_{B,e(L)} + 0} = 1 \qquad \square$$

A.9. Proof of Theorem 6

Proof. If any evidence supports proposition θ, we can always leave that evidence to be last combined without changing the final result due to the commutative property of the orthogonal sum operation. So we can have $p_{\theta,L} > 0$ from the condition of Theorem 6.

In the proof of Theorem 5, we showed that $\hat{m}_{P(\Theta),e(i)} > 0$ for any $i = 2,\ldots,L$ if Eq. (45) holds. From Eq. (25), we have $m_{\theta,L} = w_L p_{\theta,L} > 0$, so $m_{P(\Theta),e(L-1)} m_{\theta,L} > 0$. Since every other term in Eq. (32) for $i = L$ is non-negative, we get $\hat{m}_{\theta,e(L)} > 0$. From Eq. (34), we finally have $p_\theta = p_{\theta,e(L)} > 0$. $\square$

A.10. Proof of Corollary 6

Proof. If the ith piece of evidence is fully reliable, we will have $r_i = 1$. Putting $r_i = 1$ for all $i = 1,\ldots,L$ into Eqs. (35), (36) and then (40) results in

$$m_{\theta,i} = m_i(\theta) = \begin{cases} p_{\theta,i} & \theta\subseteq\Theta, \theta\neq\emptyset \\ 0 & \theta=\emptyset \\ 0 & \theta = P(\Theta)\end{cases} \quad \text{with } 0\leqslant w_i\leqslant 1 \tag{A.10.1}$$

$$\hat{m}_{\theta,e(2)} = \sum_{B\cap C=\theta} m_{B,1}m_{C,2} = \sum_{B\cap C=\theta} p_{B,1}p_{C,2} \quad \forall\theta\subseteq\Theta \tag{A.10.2}$$

For the orthogonal sum of two pieces of independent evidence, we always have

$$\sum_{\substack{B\cap C=D\\ D\subseteq\Theta}} p_{B,1}p_{C,2} + \sum_{B\cap C=\emptyset} p_{B,1}p_{C,2} = \sum_{B\cap C\neq\emptyset} p_{B,1}p_{C,2} + \sum_{B\cap C=\emptyset} p_{B,1}p_{C,2} = 1 \tag{A.10.3}$$

Putting Eqs. (A.10.1), (A.10.2) and (A.10.3) into Eq. (29) leads to

$$p_{\theta,e(2)} = \begin{cases} 0 & \theta=\emptyset \\ \frac{\sum_{B\cap C=\theta} m_{B,1}m_{C,2}}{\sum_{\substack{B\cap C=D\\ D\subseteq\Theta}} m_{B,1}m_{C,2}} = \frac{\sum_{B\cap C=\theta} p_{B,1}p_{C,2}}{\sum_{B\cap C\neq\emptyset} p_{B,1}p_{C,2}} = \frac{\sum_{B\cap C=\theta} p_{B,1}p_{C,2}}{1-\sum_{B\cap C=\emptyset} p_{B,1}p_{C,2}} & \theta\subseteq\Theta, \theta\neq\emptyset\end{cases} \tag{A.10.4}$$

$p_{\theta,e(2)}$ generated by Eq. (A.10.4) is the same as $m(\theta)$ generated by Eq. (5) given that $m_i(\theta) = p_{\theta,i}$ for any i and $\theta\subseteq\Theta$. $\square$

References

[1] M.H. Bazerman, Judgment in Managerial Decision Making, Harvard University Press, 2006.
[2] M. Beynon, DS/AHP method – A mathematical analysis, including an understanding of uncertainty, Eur. J. Oper. Res. 140 (1) (2002) 148–164.
[3] B.G. Buchanan, E.H. Shortliffe, Rule-Based Expert Systems – The MYCIN Experiments of the Stanford Heuristic Programming Project, Addison–Wesley, Reading, MA, 1984.
[4] A.P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[5] A.P. Dempster, A generalization of Bayesian inference, J. R. Stat. Soc. Ser. B 30 (1968) 205–247.
[6] T. Denoeux, L.M. Zouhal, Handling possibilistic labels in pattern classification using evidential reasoning, Fuzzy Sets Syst. 122 (3) (2001) 409–424.
[7] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artif. Intell. 172 (2008) 234–264.
[8] T. Denoeux, Z. Younes, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, Artif. Intell. 174 (7–8) (2010) 479–499.
[9] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Comput. Intell. 4 (1988) 244–264.
[10] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, Fuzzy Sets Syst. 42 (1991) 87–101.
[11] D. Dubois, H. Prade, La fusion d'informations imprécises, Trait. Signal 11 (6) (1994) 447–458.
[12] I.N. Durbach, An empirical test of the evidential reasoning approach's synthesis axioms, Expert Syst. Appl. 39 (2012) 11048–11054.
[13] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, Int. J. Inf. Secur. 9 (2002) 90–95.
[14] M.C. Florea, A.L. Jousselme, É. Bossé, D. Grenier, Robust combination rules for evidence theory, Inf. Fusion 10 (2009) 183–197.
[15] J. Gordon, E.H. Shortliffe, A method for managing evidential reasoning in a hierarchical hypothesis space, Artif. Intell. 26 (3) (1985) 323–357.
[16] R. Haenni, Shedding new light on Zadeh's criticism of Dempster's rule of combination, in: The 7th International Conference on Information Fusion (FUSION), 2005.
[17] V.N. Huynh, Y. Nakamori, T.B. Ho, T. Murai, Multiple-attribute decision making under uncertainty: the evidential reasoning approach revisited, IEEE Trans. Syst. Man Cybern., Part A 36 (4) (2006) 804–822.
[18] D. Kahneman, Thinking, Fast and Slow, Farrar, Straus and Giroux, 1st edition, Penguin Publisher, 2011.
[19] J.H. Kim, J. Pearl, A computational model for causal and diagnostic reasoning in inference systems, in: IJCAI'83 Proceedings of the Eighth International Joint Conference on Artificial Intelligence, vol. 1, 1983, pp. 190–193.
[20] C.H. Lee, A comparison of 2 evidential reasoning schemes, Artif. Intell. 35 (1) (1988) 127–134.
[21] E. Lefevre, O. Colot, P. Vannoorenberghe, Belief function combination and conflict management, Inf. Fusion 3 (2002) 149–162.

[22] S. McClean, B. Scotney, Using evidence theory for knowledge discovery and extraction in distributed databases, Int. J. Intell. Syst. 12 (1997) 763–776.
[23] C.K. Murphy, Combining belief functions when evidence conflicts, Decis. Support Syst. 29 (2000) 1–9.
[24] J. Pearl, On evidential reasoning in a hierarchy of hypotheses, Artif. Intell. 28 (1) (1986) 9–15.
[25] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[26] G. Shafer, J. Pearl, Readings in Uncertain Reasoning, Morgan Kaufmann Publishers, Inc., San Mateo, California, 1990.
[27] F. Smarandache, J. Dezert, Proportional conflict redistribution rules for information fusion, in: Florentin Smarandache, Jean Dezert (Eds.), Advances and Applications of DSmT for Information Fusion, vol. 2, American Research Press, Rehoboth, 2006, pp. 3–68.
[28] F. Smarandache, J. Dezert, J.M. Tacnet, Fusion of sources of evidence with different importances and reliabilities, in: The 2010 13th IEEE Conference on Information Fusion (FUSION), 2010, pp. 1–8.
[29] P.H. Smets, The combination of evidence in the transferable belief model, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 447–458.
[30] P.H. Smets, Analyzing the combination of conflicting belief functions, Inf. Fusion 8 (4) (2007) 387–412.
[31] F. Tonon, Using random set theory to propagate epistemic uncertainty through a mechanical system, Reliab. Eng. Syst. Saf. 85 (2004) 169–181.
[32] Y.M. Wang, J.B. Yang, D.L. Xu, The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees, Eur. J. Oper. Res. 175 (1) (2006) 35–66.
[33] D.L. Xu, J.B. Yang, Y.M. Wang, The ER approach for multi-attribute decision analysis under interval uncertainties, Eur. J. Oper. Res. 174 (3) (2006) 1914–1943.
[34] D.L. Xu, G. McCarthy, J.B. Yang, Intelligent decision system and its application in business innovative capability assessment, Decis. Support Syst. 42 (2006) 664–673.
[35] D.L. Xu, Assessment of nuclear waste repository options using the ER approach, Int. J. Inf. Technol. Decis. Mak. 8 (3) (2009) 581–607.
[36] R.R. Yager, On the Dempster–Shafer framework and new combination rules, Inf. Sci. 41 (1987) 93–137.
[37] R.R. Yager, Decision making using minimization of regret, Int. J. Approx. Reason. 36 (2004) 109–128.
[38] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple attribute decision making with uncertainty, IEEE Trans. Syst. Man Cybern. 24 (1) (1994) 1–18.
[39] J.B. Yang, P. Sen, A general multi-level evaluation process for hybrid MADM with uncertainty, IEEE Trans. Syst. Man Cybern. 24 (10) (1994) 1458–1473.
[40] J.B. Yang, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, Eur. J. Oper. Res. 131 (2001) 31–61.
[41] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 32 (3) (2002) 289–304.
[42] J.B. Yang, Y.M. Wang, D.L. Xu, The Evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties, Eur. J. Oper. Res. 171 (1) (2006) 309–343.
[43] J.B. Yang, D.L. Xu, X.L. Xie, A.K. Maddulapalli, Multicriteria evidential reasoning decision modelling and analysis – prioritising voices of customer, J. Oper. Res. Soc. 62 (2011) 1638–1654.
[44] Z.J. Zhang, J.B. Yang, D.L. Xu, A hierarchical analysis model for multiobjective decision making, in: Analysis, Design and Evaluation of Man-Machine System 1989, Pergamon, Oxford, UK, 1990, pp. 13–18.
[45] L.A. Zadeh, On the validity of Dempster's rule of combination of evidence, Technical Report 79/24, University of California, Berkeley, 1979.
[46] L.A. Zadeh, A mathematical theory of evidence, AI Mag. 55 (1984) 81–83.
[47] L.A. Zadeh, A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination, AI Mag. 7 (1986) 85–90.