

ARTICLE

# Credence and Belief: Distance- and Utility-Based Approaches

Minkyung Wang<sup>1</sup> and Chisu Kim<sup>2</sup>

<sup>1</sup>Munich Center for Mathematical Philosophy and <sup>2</sup>Independent researcher

**Corresponding author:** Minkyung Wang; Email: [minkyungwang@gmail.com](mailto:minkyungwang@gmail.com)

(Received 22 August 2022; revised 26 August 2023; accepted 26 February 2024; first published online 06 March 2024)

## Abstract

This paper investigates the question of how subjective probability should relate to binary belief. We propose new distance minimization methods, and develop epistemic decision-theoretic accounts. Both approaches can be shown to get “close” to the truth: the first one by getting “close” to a given probability, and the second by getting expectedly “close” to the truth. More specifically, we study distance minimization with a refined notion of Bregman divergence and expected utility maximization with strict proper scores. Our main results reveal that the two ways to get “close” to the truth can coincide.

## 1. Introduction

Belief can be modeled using probability functions or binary beliefs, and belief binarization investigates how they ought to relate to each other. This paper proposes new methods for rational agents with subjective probabilities to determine which propositions to believe. Most of the belief binarization literature has focused on threshold-based methods, which associate beliefs with high probabilities. The most typical one is the *Lockean thesis*, which stipulates that we ought to believe propositions with probabilities above a given threshold. However, it is well known that this approach leads to the *lottery paradox*, which underscores the tension between the proposition-wise independence norm—whether to believe a proposition should depend only on the probability of the proposition—and the rationality norm for binary belief, encompassing consistency and deductive closure. To elude the paradox, we may choose thresholds depending on the input probability functions so that only rational binary beliefs are generated, as in the stability theory of belief (Leitgeb 2017). Alternatively, thresholds could be applied not to propositions’ probabilities but rather to worlds’ probabilities, as in the tracking theory of belief (Lin and Kelly 2012a) and the normality theory of belief (Goodman and Salow 2023). However, the focus still remains on the high probability associated with each proposition or each world.

Our argument in favor of new belief binarization rules commences with a critique of the existing proposition-wise and world-wise threshold-based rules. We contend that the intuitive appeal of proposition-wise thresholds diminishes as the logical interconnection of propositions increases, as the lottery paradox already shows. The stability theory of belief also demonstrates that the logical interconnection leads to putting serious restrictions on the possible values of event-wise thresholds. The perspective of world-wise threshold rules might permit binary beliefs to deviate excessively from credence, allowing for believing propositions with a probability lower than half. Thus, we will neither collect propositions with high probability to determine a belief set (the set of believed propositions), nor collect worlds with high probability to determine a belief core whose supersets constitute the belief set. In this paper, we will develop a holistic way to determine consistent and deductively closed binary beliefs. To this end, let us focus on two related yet different aspects of the relationship between binary belief and credence.

First, binary belief might be a simplification or approximation of credence (Leitgeb 2014). This perspective becomes evident when we take probabilistic beliefs to be more refined and informative doxastic states than binary beliefs. To explicate this interpretation, we need to find a reasonable similarity measure to assess the disparity between binary beliefs and credences. By employing this measure, we can select the most similar binary beliefs to a probability function. Second, binary beliefs could be evaluated for their accuracy in light of credence. This approach is well justified, given that truth tracking stands as one of the fundamental epistemic goals. To explicate the second aspect, we need to identify a reasonable epistemic score or utility function to evaluate the epistemic performance of binary beliefs. Additionally, we should opt for a well-justified decision-theoretic principle for selecting binary beliefs in accordance with credence. The most conventional one in the decision-theoretic context is expected utility maximization (Levi 1967; Greaves and Wallace 2006; Oddie 1997). We refer to the first belief binarization methods as “distance minimization rules” (DM rules), and the second as “expected (epistemic) utility maximization rules” (EUM rules).

In this study, we first provide general forms of the DM and EUM rules, and then investigate desirable properties. The most interesting question would be the following: Is it possible to have belief binarization methods that serve as both DM and EUM rules simultaneously? This question holds philosophical significance because one might argue that to find binary beliefs similar to a credence and to track accurately the truth in light of credence are two distinctly challenging objectives to achieve simultaneously. Our main results demonstrate that the two objectives can coincide. Indeed, we will devise a specific approach for the DM rules utilizing Bregman divergences and the EUM rules employing strictly proper scoring functions, and we will show that both rules can be represented by each other.

Our approaches are situated within the belief binarization literature in the following manner. First of all, distance-based approaches have been rarely discussed in the belief binarization literature; Chandler (2013) considers only the distance from probability functions to worlds, rather than addressing binary beliefs. Regarding expected utility maximization, our EUM rules differ from the accuracy-first approach to the belief binarization problem in Dorst (2019), which does not support the deductive closure of binary beliefs. Since we believe that the most natural solution to

the lottery paradox should adhere to the logical closure of belief, our paper presupposes that all binary beliefs ought to be deductively closed, and thus, we do not take the veritistic norm as the only norm from which other rationality norms should be generated. In this regard, our theory is in line with Hempel (1960), Levi (1967), and (Leitgeb 2017, ch. 5). Lastly, to the best of our knowledge, there have been no existing studies on the relation between distance-based belief binarization methods and expected epistemic utility maximization approaches, although the relation between Bregman divergences and proper scoring rules has been extensively studied for decades (Gneiting and Raftery 2007) and employed in epistemic decision theory (Predd et al. 2009).

The rest of this paper is organized as follows: Section 2 provides an informal presentation of our main ideas. Section 3 introduces our formal setting for this paper. Section 4 explicates the DM rules and characterizes them. Moreover, it suggests the refined definition of Bregman divergence and employs it to prove that the DM rules with Bregman divergences (DM(Bregman)) can be interpreted as expected distance minimization. Section 5 formulates the EUM rules with strictly proper scores and proves that these can be represented by DM(Bregman). Section 6 concludes the paper with some discussion points.

## 2. Distance- and utility-based belief binarization

This section is devoted to an informal presentation of our central ideas, which will be spelled out with mathematical details in the other sections. We start by motivating the DM rules and EUM rules.

### *DM and EUM rules*

The DM rules can be naturally motivated by the necessity of achieving a seamless integration between our quantitative and qualitative belief states. When binary beliefs significantly deviate from the credal state, it would be difficult for an agent to organize their beliefs coherently and to guide their actions consistently. Thus, it is beneficial to facilitate a coherent linkage between credence and belief by seeking the most similar binary belief corresponding to credence. However, harmonious integration of different types of beliefs is not the sole goal of our epistemic life. So, let us motivate DM rules with respect to the truth-tracking norm.

Binary beliefs aim at the truth. Thus, our objective is to provide belief binarization rules that track the truth well—the closer to the truth, the better the rule. However, perfectly rational agents in our framework might lack the knowledge of the actual truth; they possess only subjective probability functions, which are rationally permissible credal states for truth tracking (Joyce 1998) and aim at the truth as well. Given that probabilistic beliefs are more fine-grained and more informative than binary beliefs, rational belief binarization methods should strive to identify binary beliefs as close to the probabilistic beliefs in question as possible. In this sense, this method tracks the truth by tracking probabilistic beliefs and thus belongs to a sort of implicit method adhering to the aiming-at-truth norm.

We can also conceive of belief binarization methods that explicitly consider the truth-tracking norm for beliefs. As our agents lack access to the truth but only to

subjective probabilities that encapsulate all internally accumulated evidence about the truth, the agent, at best, can expect her binary beliefs to be accurate based upon her credal state. Arguably, this method can be viewed as a sort of explicit method to track the truth. In epistemic decision theory, minimizing the expected inaccuracy of credence has been employed to justify epistemic norms (Greaves and Wallace 2006; Oddie 1997). We apply the minimization of the expected inaccuracy of belief, as decision rules, directly to the belief binarization problem (Levi 1967). In conclusion, to get “close” to the truth, the best approaches for an agent with a probability function to determine what to believe would involve either (i) the ways to get “close” to the probability function, or (ii) those to *expectedly* get “close” to the truth. The former is the DM rules, and the latter is the EUM rules.

### *Distance between binary beliefs and credences*

The DM rules determine binary beliefs that minimize the distance from a given probability function  $P$ . To measure the distance from a probability function to a belief set, we will utilize divergences between probability functions. However, a belief set is not a probability function. Our key concept involves linking a belief set  $Bel$  with a probability function: the uniform distribution  $U(B)$  over the belief core  $B$  of  $Bel$  (the conjunction of all believed propositions), which is guaranteed to be believed by the assumption of deductive closure. Building on this, we can assess the discrepancy between a probability function and a belief set  $Bel$  by measuring the divergence between the probability function and  $U(B)$ . Accordingly, the distance from  $U(B)$  to  $Bel$  becomes zero, and the DM rules map  $U(B)$  to  $Bel$ . This does not mean that  $U(B)$  is the probabilistic representation of a rational belief set  $Bel$ . Following the process of belief binarization, we might generate possibly infinitely many probability functions corresponding to specific binary beliefs  $Bel$ . The probability function  $U(B)$  can be interpreted just as the most representative credal state corresponding to  $Bel$ .

To justify our methodology, we need to explain why  $U(B)$  is the most natural probability function generating the set  $Bel$  of all the supersets of  $B$  as the belief set. We suggest an epistemic principle, termed the suspension principle, which posits that belief binarization rules ought to map  $U(B)$  to  $Bel$ . Let's delve into the plausibility of this principle. In cases where  $B$  is the singleton set  $\{w\}$  of a world  $w$ ,  $U(B)$  represents the probabilistically certain belief that  $w$  is the actual world. Thus, the resulting belief set  $Bel$  should indeed be the set of all supersets of  $\{w\}$ . When  $B$  is not a singleton,  $U(B)$  represents the probabilistically certain belief that any world lying outside of  $B$  is not the actual world. Therefore, we can exclude those worlds. Moreover, since  $U(B)$  is uniform over  $B$ , it represents that the agent lacks an opinion regarding which world in  $B$  is the actual one. Thus, the agent should suspend judgment about whether the actual world resides in any strict subset of  $B$ , and thereby  $B$  should be the belief core. And since  $B$  has a probability of 1, all supersets of  $B$  should be believed. By contrast, since probability functions other than  $U(B)$  break the symmetry—regardless of how slight a perturbation may be—we cannot use this symmetry consideration for them.

Based on this framework, our characterization theorem of DM rules establishes that upholding the suspension principle amounts to employing our DM rules. It is worth noting that the suspension principle is not demanding and serves as a minimal requirement when permitting the suspension of judgments. This condition is lenient enough to embrace almost all threshold-based rules in the relevant literature.

Nevertheless, our DM rules do not encompass every conceivable distance-based belief binarization approach, and there might be other ways to measure distance for belief binarization. But our characterization theorem of DM rules implies that if alternative distance-based rules satisfy the suspension principle, they are subsumed under our one. And the violation of the suspension principle demonstrates a significant drawback inherent in them. In fact, it can be easily checked that the Hamming rule presented by Dietrich and List (2021) violates the suspension principle.

### *DM(Bregman) and EUM(SP) rules*

Now the problem is reduced to how one can measure the distance from a probability function to the uniform distribution on a belief core, i.e., a non-empty set of possible worlds. Many kinds of distance measures exist. Could there be some rationality norm to guide one in selecting distance measures? Our first main theorem will answer that question: if we employ a Bregman divergence for the DM rules, which will be called *DM(Bregman)*, these rules can be represented by the expected utility maximization procedure, where the disutility is the Bregman divergence from the truth. If we agree that maximizing expected utility is a standard strategy to choose binary beliefs given a credal state, this theorem reveals a good reason why we should seriously consider distance minimization methods.<sup>1</sup>

Let us elaborate further on this. Within our framework, the disparity between a world  $w$  and a rational belief set  $Bel$  can be assessed by measuring the divergence between the omniscient probability at  $w$  and  $U(B)$ . Furthermore, this divergence can be regarded as an epistemic inaccuracy of a belief set  $Bel$  at a world  $w$ . Since the most relevant feature of Bregman divergences concerning our study is that minimizing a Bregman divergence from  $P$  is tantamount to minimizing the expected Bregman divergence from the true world in light of  $P$ , any *DM(Bregman)* rule can also be viewed as an EUM rule. Additionally, the epistemic inaccuracy of a belief set at a world, as determined by a Bregman divergence, is a strictly proper score. So we now shift our focus to the EUM rules with strictly proper scores, called *EUM(SP)*.

Strictly proper scores are prominently featured inaccuracy measures in epistemic decision theory and widely adopted as utilities in various other domains, such as probabilistic forecasting and belief elicitation. Strict propriety requires that the expected utility of a probability function according to  $P$  is uniquely maximized at  $P$ . Therefore, to attain maximal expected utility, we must opt for  $P$  when it is provided, leading to the suspension principle in our framework. This observation implies that *EUM(SP)* rules can be regarded as DM rules. Moreover, building upon the fact that expected strictly proper scores generate Bregman divergences, our second main theorem demonstrates that *EUM(SP)* rules are *DM(Bregman)* rules.

Figure 1 depicts the relationships between the aforementioned rules. Our findings underscore that *EUM(SP)* rules, or equivalently *DM(Bregman)* rules, are the ways to get close to the probability and to expectedly get close to the truth at the same time. Both DM and EUM rules possess their own merits: the former can build a harmonious relationship between credence and belief by seeking the most similar binary beliefs to

<sup>1</sup> In Wang (n.d.), we showed that some other existing belief binarization methods, such as Leitgeb's stability theory of belief and Lin-Kelly's tracking theory of belief, cannot be represented by any maximizing expected utility procedure.

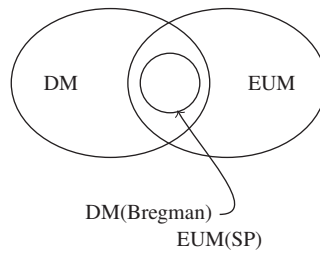


Figure 1. DM and EUM rules.

credence; the latter, on the other hand, guide in finding binary beliefs that approximate the truth. So, it would be beneficial to identify a class of rational belief binarization rules satisfying two principles simultaneously. Indeed, our representation theorems substantiate the existence of an extensive range of rational belief binarization rules that uphold these dual principles.

### Generalization

We aim to develop our distance- and utility-based approaches within a comprehensive framework that can accommodate a wide range of scenarios. In the literature concerning belief binarization, probability functions are typically represented by points within a probability simplex. In our work, we not only utilize the probability simplex for representing probability functions, but we also incorporate De Finetti's coherent polytope. In the former approach, each world's probability (we will assume worlds to be finite) will be relevant, while the probabilities of some focused events, called an agenda, will be relevant in the latter approach. The latter approach exhibits greater generality than the former. If all the singleton sets of worlds constitute the agenda, then the latter approach turns into the former. This broadened perspective aligns our framework with the literature on epistemic decision theory or belief aggregation (Pettigrew 2015; Dietrich and List 2017a) and provides a new viewpoint for the belief binarization problem. In this framework, we can consider an agenda for the belief binarization problem. The agenda does not need to form an algebra. Thus, when we apply our DM rules in this framework, we do not need to have all the probabilities of propositions to determine what to believe. Even when we have the probabilities of only the propositions in the agenda, we can employ the DM rules, while other belief binarization theories in the literature cannot be applied in this case. Even if a probability function is given on an algebra, we may exclusively focus on certain basic propositions or premises. Then, our framework allows us to model a form of premise-based belief binarization, similar to the case of generalized probabilistic opinion pooling (Dietrich and List 2017b). Moreover, our methods can deal with disutility functions that depend solely on the probabilities of the focused events, although acquiring the probabilities associated with possible worlds is essential for calculating expected utility for the EUM rules.

Further challenges in this study are twofold. Firstly, we aim to refine conventional definitions of Bregman divergence. Specifically, we intend to adapt Bregman divergence to address belief binarization problems, allowing for infinite divergence within certain boundary regions of a probability simplex or De Finetti's coherent polytope. Secondly, we do not impose the constraint of proposition-wise additivity on

scoring rules. The aforementioned rules and claims will be elaborated and proved based on these technical settings. The rest of this paper elucidates these ideas with detailed mathematical explanations.

### 3. Formal setting

We let  $W$  be a finite non-empty set of possible worlds and denote by  $\mathcal{P}(W)$  the powerset algebra of  $W$ . A *probability function* on  $\mathcal{P}(W)$  is a function  $P : \mathcal{P}(W) \rightarrow [0, 1]$  satisfying the probability axioms, and we write  $\mathbb{P}(W)$  for the set of all probability functions on  $\mathcal{P}(W)$ . A *binary belief function* is a function  $Bel : \mathcal{P}(W) \rightarrow \{0, 1\}$ , and by abusing the notation, we also use  $Bel$  to denote the set of all believed events, called the *belief set*. A *belief binarization rule* (BR)  $G$  is defined to be a function that takes as input any probability function  $P$  in  $\mathbb{P}(W)$  and outputs the set of some binary beliefs. Note that  $G$  can be seen as a *correspondence* in the sense that  $G$  might output multiple binary beliefs. One could combine this with some tie-breaking rule to choose only one binary belief, if it is needed.

We assume that binarization rules  $G$  are *rational* in the sense that every resulting binary belief  $Bel$  is consistent (the belief set does not entail a contradiction) and deductively closed (the belief set contains all its logical consequences). Formally, they are defined as follows: (i)  $Bel$  is *consistent* if the intersection of all believed events is not empty, i.e.,  $\cap Bel \neq \emptyset$ ; (ii)  $Bel$  is *deductively closed* if  $Bel$  contains  $W$  and is closed under intersection and superset. It is well known that  $Bel$  is rational if and only if (iff)  $Bel$  has a non-empty belief core  $B (= \cap Bel)$  whose supersets are exactly the believed events. From now on, we regard a rational binary belief in  $G(P)$  as a *non-empty belief core*, i.e., a non-empty subset  $B$  of  $W$ .

## 4. Distance minimizing binarization rules

### 4.1. DM rules and the suspension principle

In this section, we define the distance minimizing binarization rules and characterize them. For this purpose, we need to measure the distance between the input of a BR  $G$ —a probability function  $P$  on  $\mathcal{P}(W)$ —and a non-empty subset  $B$  of  $W$ . Our first main idea is to employ a *divergence* on a convex subset of  $\mathbb{R}^m$  for some  $m \in \mathbb{N}$ .<sup>2</sup> To this end, we need to represent probability functions  $P$  and subsets  $B$  in  $\mathbb{R}^m$ . For probability functions, we could deploy some typical methods to represent probability functions in  $\mathbb{R}^m$ . However, how can we represent a subset  $B$  in  $\mathbb{R}^m$ ? Our main idea is to associate it with the *uniform distribution*  $U(B)$  on  $B$ —the probability distribution that assigns  $1/|B|$  to each world in  $B$  and 0 to other worlds. It is plausible because when the input of a BR  $G$  is  $U(B)$ ,  $B$  must be the most natural belief binarization result, and thus we want to set the distance between  $U(B)$  and  $B$  equal to 0.

### Representations of probabilities

The remaining aspect of measuring distance involves addressing how to represent probability functions in  $\mathbb{R}^m$ . Two approaches have been adopted in this regard. In certain contexts, such as belief binarization theories or statistics, probabilities of

<sup>2</sup> We call a function  $d : X \times X \rightarrow [0, \infty]$  a divergence on a convex set  $X \subseteq \mathbb{R}^m$  when  $d(x, y) = 0$  for all  $x, y \in X$ , where the equality holds iff  $x = y$ .

worlds are employed. In other contexts, like belief aggregation or epistemic decision theory, a set of propositions is initially given, with only their respective probabilities being pertinent. We incorporate both of these approaches in our discussion. Let  $W = \{w_1, \dots, w_n\}$ . Our first approach is to represent a probability function  $P$  by a point  $p$  in  $\mathbb{R}^{|W|}$  such that

$$p = (p_{w_1}, \dots, p_{w_n}) = (P(w_1), \dots, P(w_n)) \in \Delta^W,$$

where  $\Delta^W$  denotes the set of all points representing probability distributions. We say that  $p$  is the *representation point of  $P$  in  $\Delta^W$* . According to this representation method, we can represent an *omniscient credence function*  $V_w$  at  $w \in W$ —assigning 1 to  $w$ —by a point  $v_w$  on  $\{0, 1\}^{|W|}$  where the  $w$ 'th coordinate is given by  $(v_w)_{w'} = V_w(w') = 1$  if  $w' = w$  and  $(v_w)_{w'} = V_w(w') = 0$  otherwise. Thus,  $\Delta^W$  is the convex hull of the representation points of all omniscient credence functions because  $p = \sum_{w \in W} P(w)v_w$ . Note that  $\Delta^W$  is nothing but the usual  $(|W| - 1)$ -dimensional probability simplex.

We now move to the second approach. We introduce a non-empty subset  $\mathcal{F} = \{A_1, \dots, A_m\} \subseteq \mathcal{P}(W)$ , and call it the *set of focused events* or an *agenda*. Even though probabilistic and binary beliefs are functions from  $\mathcal{P}(W)$ , there can be some situations where we are only interested in the focused events in  $\mathcal{F}$ . In this case, we can represent probability functions in  $\mathbb{R}^{|\mathcal{F}|}$  and measure distances between them in this space. Note that probabilistic and binary beliefs are assumed to be functions not from  $\mathcal{F}$  but from  $\mathcal{P}(W)$  even in this approach as well, which can be relaxed later.<sup>3</sup> We represent a probability function  $P$  by a point  $p$  in  $\mathbb{R}^{|\mathcal{F}|}$  such that

$$p = (p_{A_1}, \dots, p_{A_m}) = (P(A_1), \dots, P(A_m)) \in \Delta^{\mathcal{F}},$$

where we denote by  $\Delta^{\mathcal{F}}$  the set of all points representing probability distributions. We say that  $p$  is the *representation point of  $P$  in  $\Delta^{\mathcal{F}}$* . In this approach, an omniscient credence function  $V_w$  at  $w \in W$  is represented by a point  $v_w$  on  $\{0, 1\}^{|\mathcal{F}|}$  where the  $A$ th coordinate  $(v_w)_A = V_w(A)$  is 1 if  $w \in A$  and 0 otherwise. Accordingly,  $\Delta^{\mathcal{F}}$  is the convex hull of  $\{v_w \in \{0, 1\}^{|\mathcal{F}|} \mid w \in W\}$ . Notice that  $\Delta^{\mathcal{F}}$  is a 0/1-polytope in  $\mathbb{R}^{|\mathcal{F}|}$  (a polytope whose vertices are on  $\{0, 1\}^{|\mathcal{F}|}$ ).

It is interesting to compare the two approaches. In the case where  $\mathcal{F} = \{\{w\} \in \mathcal{P}(W) \mid w \in W\}$ , both ways are the same. If  $|\mathcal{F}| < |W|$ , one point in  $\Delta^{\mathcal{F}}$  represents several distinct probability distributions, i.e., a probability distribution is not uniquely determined by a point: a point in  $\Delta^{\mathcal{F}}$  represents a convex set of probability functions. This is because if  $P, P'$  have the same representation  $p$ , then for all  $A \in \mathcal{F}$  we have  $P(A) = P'(A) = \alpha P(A) + (1 - \alpha)P'(A)$  for all  $\alpha \in [0, 1]$ , which means that any linear combination of them has the same representation. Therefore, we can regard a point in  $\mathcal{F}$  as a convex set of probability functions. Note that many definitions, theorems, and statements in this paper will be formulated using not only the representations in  $\Delta^W$  but also the ones in  $\Delta^{\mathcal{F}}$ . To express this, we will use  $\Delta^M$ . Hence  $M$  will be considered to be  $W$  or  $\mathcal{F}$  throughout this paper.

<sup>3</sup> We need this condition to well define expected epistemic utility in EUM rules where we need to calculate a probability of a singleton world, unless we have IER (invariant expectation under the same representation in Definition 8).



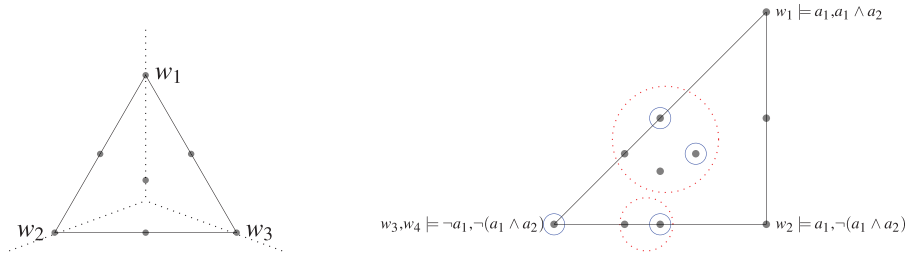


Figure 2. Belief binarization problems.

Representations of belief cores

Now let us turn to the representations of the uniform distribution  $U(B)$  on a non-empty belief core  $B \in \mathcal{P}(W) \setminus \{\emptyset\}$ . We will use the small letter  $b$  to refer to the representation point of  $U(B)$  in  $\Delta^M$  and call it the corresponding point of  $B$  in  $\Delta^M$ . Moreover, we denote by  $U^M$  the set of all corresponding points of non-empty belief cores, i.e.,  $U^M = \{b \in \Delta^M \mid \emptyset \neq B \subseteq W\}$ . With this in place, we will, hereafter, let a BR  $G$  be a correspondence from  $\mathbb{P}(W)$  to  $U^M$ . In other words,  $G(P)$  refers to a set of points  $b$  in  $U^M$ . Since we address binarization rules based on distance minimization or expected utility maximization, a BR  $G$  will take the form of  $\text{argmin}_b g(P, b)$ , which is the set of all points  $b$  in  $U^M$  that minimize  $g(P, b)$ , where  $g$  is an extended real-valued function. Note that the uniform distribution on a belief core might not be uniquely determined by a point in  $\Delta^{\mathcal{F}}$  as explained above, which can lead to the under-determination of a belief core. However, even if a point that corresponds to several belief cores is selected, one could combine it with a tie-breaking rule. Moreover, the following lemma shows that even though different belief cores give us different belief sets, if their corresponding points in  $\Delta^{\mathcal{F}}$  are the same, then they yield the same result concerning any event in  $\mathcal{F}$ .

**Lemma 1** (Invariance under the same output representation). *Let  $B$  and  $B'$  be non-empty subsets of  $W$ . If the corresponding points in  $\Delta^{\mathcal{F}}$  are the same, i.e.,  $b = b'$ , then for all  $A \in \mathcal{F}$ , (i)  $B \subseteq A$  iff  $b_A = 1$  and  $B \subseteq A^c$  iff  $b_A = 0$ , and thus (ii)  $B \subseteq A$  iff  $B' \subseteq A$  and  $B \subseteq A^c$  iff  $B' \subseteq A^c$ , where  $A^c$  is the complement of  $A$ .*

The statement (i) gives us a geometrical intuition. If  $b_A = 1$ , then  $A$  is believed; if  $b_A = 0$ , then  $A^c$  is believed; if  $b_A \neq 0, 1$ , then neither  $A$  nor  $A^c$  is believed. This explains why binary beliefs on the focused events in  $\mathcal{F}$  are invariant under the same output representation (IOR).

Let us now illustrate some examples of belief binarization problems. Consider the following two binarization problems: (1)  $W = \{w_1, w_2, w_3\}$ , and (2)  $\mathcal{F} = \{[a_1], [a_1 \wedge a_2]\}$  and  $W = \{w_1 (\models a_1, a_2), w_2 (\models a_1, \neg a_2), w_3 (\models \neg a_1, a_2), w_4 (\models \neg a_1, \neg a_2)\}$ , where  $a_1$  and  $a_2$  are atomic formulas in the standard propositional logic, and for any formula  $\phi$ ,  $[\phi]$  is the set of all valuations under which  $\phi$  holds. We write  $w \models \phi_1, \phi_2$  when  $\phi_1$  and  $\phi_2$  hold under the valuation  $w$ . Figure 2 depicts (1) and (2) in  $\Delta^W$  and  $\Delta^{\mathcal{F}}$ , respectively. Each point represents the uniform distribution on a belief core. In the right panel, some of them represent several uniform distributions. The solid circles express that there are

extra uniform distributions. Note that the points that are surrounded by a dotted circle give us the same beliefs/disbeliefs of the focused events in  $\mathcal{F}$  by Lemma.

### *DM rules and the suspension principle*

Now we deploy a divergence  $d$  in  $\Delta^M$  and the representation methods discussed above in order to formulate distance minimizing binarization rules (DM rules) as follows.

**Definition 1** (Distance minimization rule (DM rule)). A BR  $G$  is a distance minimization rule (DM rule) in  $\Delta^M$  iff there is a divergence  $d$  in  $\Delta^M$  such that, for all  $P \in \mathbb{P}(W)$ ,  $G(P) = \operatorname{argmin}_b d(p, b)$ , where  $p$  is the representation of  $P$  in  $\Delta^M$ .

So a DM rule with a divergence  $d$  in  $\Delta^M$  is a correspondence that takes as input any probability function  $P$  and outputs the points  $b$  in  $U^M$  that minimize the distance from  $p$ . The following theorem states that the DM rules are characterized by an epistemic principle, what we call the suspension principle. Here is the formal definition of the suspension principle.

**Definition 2** (Suspension principle). A BR  $G$  satisfies the suspension principle in  $\Delta^M$  iff, for all  $B \in \mathcal{P}(W) \setminus \{\emptyset\}$ ,  $G(U(B)) = \{b\}$ , where  $b$  is the representation point of  $U(B)$  in  $\Delta^M$ .

This means that if  $P$  is a uniform distribution on certain worlds, then binarization rules should result in the belief core that consists of those worlds. For the characterization theorem, we will need the following condition, which provides some control over the cases where points in  $\Delta^M$  can represent several probability distributions.

**Definition 3** (Invariance under the same input representation). A BR  $G$  satisfies invariance under the same input representation (IIR) in  $\Delta^M$  iff, for all  $P, P' \in \mathbb{P}(W)$  with the same representation in  $\Delta^M$ , i.e.,  $p = p' (\in \Delta^M)$ ,  $G(P) = G(P')$ , where  $p$  and  $p'$  are the representation points of  $P$  and  $P'$  in  $\Delta^M$ , respectively.

If  $G$  satisfies IIR in  $\Delta^{\mathcal{F}}$ , then the binarization results depend only on the probabilities of the focused events in  $\mathcal{F}$ . Thus, the representation point  $p$  plays the role of the input of  $G$ . This amounts to dealing with a generalized agenda, which does not need to be an algebra, and *probabilistically coherent beliefs*—a function extendable to a probability function on the algebra generated by the generalized agenda. Let us compare this with the invariance under the output representation in Lemma 1. IIR means that two probability functions with the same representation input point give us the same output point, which is associated with several belief cores. On the other hand, IOR in Lemma 1 shows that binary belief functions corresponding to an output point give us the same beliefs or disbeliefs about the focused events in  $\mathcal{F}$ . The following theorem says that the suspension principle characterizes the DM rules if  $G$  satisfies IIR.

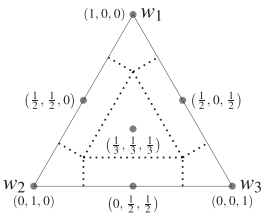


Figure 3. DM(SE) when  $W = \{w_1, w_2, w_3\}$ .

**Theorem 1** (Characterization of DM rule). *A BR  $G$  is a DM rule in  $\Delta^M$  iff (i)  $G$  satisfies IIR in  $\Delta^M$ , and (ii)  $G$  satisfies the suspension principle.*

We remark that DM rules in  $\Delta^W$  always satisfy IIR. Thus, a BR  $G$  is a DM rule in  $\Delta^W$  iff  $G$  satisfies the suspension principle. From this theorem, it can be easily checked that the threshold-based rules in Leitgeb (2017), Lin and Kelly (2012b), and Goodman and Salow (2023) can also be seen as a DM rule in  $\Delta^W$ .

The most natural DM rule can be given by using the squared Euclidean distance  $D_{SE}$  in  $\Delta^W$ .<sup>4</sup> We call this the DM(SE) rule in  $\Delta^W$ . Figure 3 illustrates DM(SE) for the case where  $|W| = 3$ . The seven dots represent the uniform distributions  $b$  of the seven belief cores. The dotted lines divide the simplex into seven regions. Each region is the preimage region  $G^{-1}(b) (= \{p \in \Delta^W \mid b \in G(P)\})$  of the point  $b$  under  $G$ . For example, consider a probability distribution  $P$  such that  $P(w_1) = 0.8$  and  $P(w_2) = 0.2$ , which can be represented by a point  $p = (0.8, 0.2, 0)$ . According to DM(SE),  $P$  should be assigned to the belief core  $\{w_1\}$  because the squared Euclidean distance between  $p$  and  $(1, 0, 0)$  is much less than the distances of  $p$  to the other six corresponding points of belief cores. Additionally, it is also easily checked that if we use the inverse Kullback–Leibler divergence (IKL)<sup>5</sup> in  $\Delta^W$ , the resulting belief binarization, called DM(IKL), is the same as the probability 1 proposal (we ought to believe a proposition  $A$  iff  $P(A) = 1$ ). In the following section, we will generalize the cases of DM(SE) and DM(IKL) to the DM rules with a Bregman divergence.

### 4.2. DM rules with Bregman divergences

#### Refined Bregman divergence

First of all, we refine the definition of Bregman divergence to make it applicable to the belief binarization problems. In the typical definition of Bregman divergence in most of the literature, the domain of the second argument is an open set—e.g., the (relative) interior of a probability simplex  $\Delta^W$  or  $\mathbb{R}^m$ —or the value of Bregman divergence cannot be infinity. For our purposes, however, we need to define it in a closed set  $\Delta^M$ , because many uniform distributions on belief cores are located on the boundary of the set. Furthermore, we should allow infinity as a possible value of divergence, because we want to embrace some asymptotically divergent distance measures like the inverse Kullback–Leibler divergence. For this reason, we need to extend the definition of Bregman divergence with *infinity* to the (relative) boundary of  $\Delta^M$ .

<sup>4</sup> Note that  $D_{SE}(p, q) := \sum_{w \in W} (P(w) - Q(w))^2$ .

<sup>5</sup> The definition of IKL is as follows:  $D_{IKL}(p, q) = \sum_{w \in W} P(w) \log(P(w)/Q(w))$  if, for all  $w$ ,  $P(w) \neq 0$  implies  $Q(w) \neq 0$ , otherwise  $\infty$ .

Considering that our definition includes infinity on the boundary, a comparison can be drawn with the works of Adamcik (2014) and Predd et al. (2009). In Adamcik (2014), Bregman divergence is defined in  $\Delta^W$  and its value can be infinity.<sup>6</sup> Our definition is more general because it embraces not only Bregman divergences in  $\Delta^W$  but also those in  $\Delta^{\mathcal{F}}$ . Turning to Predd et al. (2009), Bregman divergence is defined in  $\Delta^{\mathcal{F}}$  and its value can be infinity. However, the additivity of Bregman divergence is presupposed. Our intention is to develop definitions and theorems in a more comprehensive manner without the assumption of additivity. Thus, our definitions and theorems work *not only in a simplex  $\Delta^W$  but also in a general 0/1-polytope  $\Delta^{\mathcal{F}}$*  dealing with not only additive but also *non-additive* divergence.

Let's elaborate on the main desired features of our proposed definition. We aim to extend the definition of finite Bregman divergence in the interior of  $\Delta^M$  to encompass the boundary, even considering infinite values. However, our intent is not to have infinite distance along the entire boundary, but rather to carefully determine where finiteness and infiniteness should be applied. For instance, take two points on the boundary that reside within the relative interior of the same lower-dimensional face. In this scenario, we want the divergence between them to remain finite, just like the divergence between two points in the interior of the 0/1-polytope. Moreover, we want the divergence to be continuous in the region where it should be finite, just as the Bregman divergence is continuous in the interior of the 0/1-polytope. On top of that, we aim to retain key properties—e.g., the relation between distance minimization and expected score maximization—and well-known instances of Bregman divergence—e.g., the squared Euclidean distance and inverse Kullback–Leibler divergence. This enables one to use the existing results about Bregman divergences. In a nutshell, our definition is designed to ensure that the behavior of divergence on the boundary mirrors that in the interior. This is reasonable in the context of the belief binarization problem because we should maintain the same belief–credence connection principles even when narrowing our attention to a specific region on the boundary by assigning zero probability to some worlds or propositions.

To this end, we provide a way to denote faces not only in the simplex  $\Delta^W$  but also in the 0/1-polytope  $\Delta^{\mathcal{F}}$ . Within  $\Delta^W$ , we can employ  $\text{Supp}(Q)$ , defined as  $\{w \in W \mid Q(w) \neq 0\}$ , to signify the set of the worlds corresponding to all vertexes of the lowest-dimensional face on which  $q$  ( $\in \Delta^W$ ) lies. We need to extend this notion to indicate the faces of  $\Delta^{\mathcal{F}}$ , where a point in  $\Delta^{\mathcal{F}}$  can represent multiple probability functions.

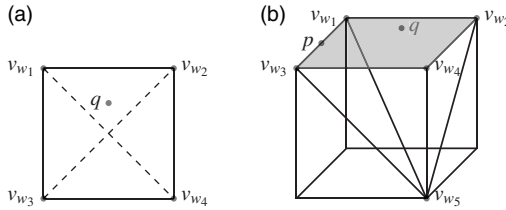
**Definition 4** (Maximal support,  $\Delta_q$  and  $\mathbb{F}_p$ ). *Let  $p, q \in \Delta^M (\subseteq \mathbb{R}^m)$ .*

(i) *The maximal support of  $q$  is defined by*

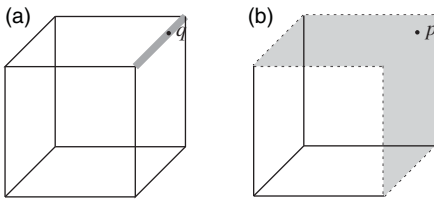
$$\text{MSupp}(q) := \bigcup_Q \text{Supp}(Q),$$

*where the  $Q$  are the probability distributions represented by  $q$ .*

<sup>6</sup> In  $\Delta^W$ , our definition looks simpler, but we can easily prove that Adamcik's definition coincides with ours except that we have a continuity condition.



**Figure 4.** In (a),  $q$  is the representation of  $Q_1, Q_2,$  and  $Q_3$  where  $\text{Supp}(Q_1) = \{w_1, w_2, w_3\}, \text{Supp}(Q_2) = \{w_1, w_2, w_4\},$  and  $\text{Supp}(Q_3) = \{w_1, w_2, w_3, w_4\}.$  Thus,  $\text{MSupp}(q) = \{w_1, w_2, w_3, w_4\}.$  In (b),  $\text{MSupp}(p) = \{w_1, w_3\} \subseteq \text{MSupp}(q) = \{w_1, w_2, w_3, w_4\}.$



**Figure 5.** When  $\Delta^M = [0, 1]^3,$   $\Delta_q$  is the thick gray line including the end points and  $\mathbb{F}_p$  is the union of the relative interior  $((0, 1)^3)$  of  $\Delta^M$  and the gray area excluding the dotted boundary.

(ii) The lowest-dimensional face on which  $q$  lies is defined by the convex hull of  $\text{MSupp}(q),$  i.e.,

$$\Delta_q := \text{Conv}(\text{MSupp}(q)) = \{x \in \Delta^M \mid \text{MSupp}(x) \subseteq \text{MSupp}(q)\}.$$

(iii) The (disjoint) union of all the relative interiors of the faces that  $p$  lies on is

$$\mathbb{F}_p := \{x \in \Delta^M \mid \text{MSupp}(p) \subseteq \text{MSupp}(x)\} = \{x \in \Delta^M \mid p \in \Delta_x\}.$$

So,  $w \in \text{MSupp}(q)$  means that there exists a probability function  $Q$  represented by  $q$  such that  $Q(w) \neq 0$ —i.e., a probability represented by  $q$  assigns to  $w$  a non-zero probability (see figure 4). Using the notion of maximal support, we can designate the lowest-dimensional face  $\Delta_q$  on which  $q$  lies, which is the convex hull of the maximal support of  $q.$  We can easily check that  $\Delta_q$  is a (sub-)0/1-polytope that is the set of the points whose maximal support is a subset of  $q$ 's maximal support. In contrast,  $\mathbb{F}_p$  is the set of the points whose maximal support is a superset of  $p$ 's maximal support (see figure 5).

Now we are ready to formulate our definition of Bregman divergence  $D.$  We modify its typical definition to the extent that  $D(p, q)$  is finite and continuous so far as  $q \in \mathbb{F}_p,$  i.e.,  $\text{MSupp}(p) \subseteq \text{MSupp}(q),$  which means that all worlds that are probabilistically possible according to some probability function represented by  $p$  are also probabilistically possible according to some probability function represented by  $q.$  Loosely speaking, we aim for Bregman divergences to remain finite as far as  $q$  does not exclude any world that  $p$  does not exclude.

**Definition 5** (Refined Bregman divergence).  $D : \Delta^M \times \Delta^M \rightarrow [0, \infty]$  is a Bregman divergence in  $\Delta^M$  iff there is a continuous, bounded, and strictly convex function  $\Phi : \Delta^M \rightarrow \mathbb{R}$  satisfying the following for all  $p, q \in \Delta^M$ :

- (i) if  $q \in \mathbb{F}_p$ , then the directional derivative  $\nabla_{p-q}\Phi(q)$  in the direction of  $p - q$  at  $q$  exists, being finite and continuous in  $q$ , and

$$D(p, q) = \Phi(p) - \Phi(q) - \nabla_{p-q}\Phi(q);$$

- (ii) otherwise,  $D(p, q) = \lim_{x \rightarrow q : x \in \mathbb{F}_p} D(p, x)$ , which exists, infinity being allowed as limits.

This definition is compared with the conventional ones of Bregman divergence in the relative interior of  $\Delta^M$ , denoted by  $\text{ri}(\Delta^M)$ , as follows. As usual, Bregman divergence is defined in terms of a convex function  $\Phi$  called a Bregman divergence generator. What distinguishes our definition from the conventional ones is parts (i) and (ii). In the conventional definitions, part (i) is applied in  $\text{ri}(\Delta^M)$ , which is the whole domain in the conventional ones, and (ii) is not needed. By contrast, we apply part (i) to the region where  $q \in \mathbb{F}_p$ , and extend this continuously, infinity being allowed as limits, to the rest of the domain.

Note that we use the directional derivative instead of the gradient in the conventional definition. This is because we need to define divergence not only in the interior but also on the boundary, where gradients are not well defined. In  $\text{ri}(\Delta^M)$ ,  $\nabla_{p-q}\Phi(q) = \nabla\Phi(q) \cdot (p - q)$  because  $\Phi$  is differentiable from part (i) by the convexity of  $\Phi$ .<sup>7</sup> In the interior of any lower-dimensional face  $\Delta_q$ , we could also say that  $\Phi|_{\text{ri}(\Delta_q)}$  (the restriction of  $\Phi$  to  $\text{ri}(\Delta_q)$ ) is differentiable in the sense that it is differentiable in the lower-dimensional space (note that the affine hull of  $\Delta_q - q$  ( $:= \{x - q | x \in \Delta_q\}$ ) is a subspace of  $\mathbb{R}^m$ ). In this sense, we could conclude that our definition of  $D(p, q)$  coincides with the conventional one not only for  $p, q \in \text{ri}(\Delta^M)$  but also for  $p, q$  in the relative interior  $\text{ri}(\Delta_q)$  of any lower-dimensional face.

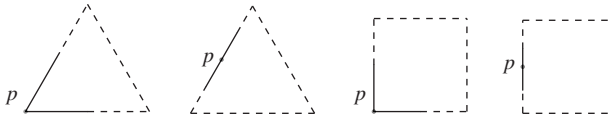
Now let us consider another way to extend the conventional definitions of Bregman divergence to the boundary to compare this with our definition. Instead of parts (i) and (ii), we could have defined Bregman divergence as follows:

- (\*) For all  $p, q \in \Delta^M$ , if  $q \in \text{ri}(\Delta^M)$ , then  $\nabla_{p-q}\Phi(q)$  exists, being finite,<sup>8</sup> and  $D(p, q) = \Phi(p) - \Phi(q) - \nabla_{p-q}\Phi(q)$ , otherwise  $D(p, q) = \lim_{x \rightarrow q : x \in \text{ri}(\Delta^M)} D(p, x)$  which exists, infinity being allowed as limits.

According to this definition, the divergence is finite and continuous in  $\text{ri}(\Delta^M)$ , and it does not guarantee finite and continuous divergence between two points in the relative interior of a face on the boundary. Figure 6 shows the problem cases that could arise if we defined Bregman divergence according to (\*). In these cases, the divergence  $D(p, q)$  can be infinite although  $q$  does not exclude any world that  $p$  does not exclude. We will see later that if we do not prevent these cases, then we cannot prove the relation between Bregman divergences and proper scores in Theorem 2.

<sup>7</sup> See Lemma 2 in the Appendix, and Rockafellar (1970, Theorem 25.2).

<sup>8</sup> In this case, since  $\Phi$  is convex,  $\nabla_{p-q}\Phi(q)$  is continuous in  $q$  (see Lemma 2 in the Appendix and Rockafellar (1970, Theorem 25.2 and Corollary 25.5.1)).



**Figure 6.** These cases can occur according to the definition of (\*). The dashed lines in each polytope represent where  $D(p, \cdot)$  is infinite.

For this reason, it is hoped that the region where Bregman divergence should be finite and continuous is extended from  $\text{ri}(\Delta^M)$  to  $\mathbb{F}_p$ .

**4.3. Representation of DM(Bregman) by expected distance minimization**

Now we employ our refined Bregman divergence for DM rules. A DM(Bregman) rule is the DM rule with a Bregman divergence  $D$  in  $\Delta^M$ , which has the following form: for all  $P \in \mathbb{P}(W)$ ,  $G(P) = \text{argmin}_b D(p, b)$ , where  $p (\in \Delta^M)$  is the representation point of  $P$ .

Now, our aim is to prove that the DM(Bregman) rule can be represented by a decision rule that minimizes expected distance from the point  $v_w (\in \Delta^M)$  corresponding to a world  $w \in W$ , which is the representation point of the omniscient credence function  $V_w$  at  $w \in W$ . (Recall that if  $\Delta^M$  is  $\Delta^W$ ,  $(v_w)_{w'} = V_w(w') = \mathbb{1}_{w=w'}$ , and if  $\Delta^M$  is  $\Delta^{\mathcal{F}}$ ,  $(v_w)_A = V_w(A) = \mathbb{1}_{w \in A}$ .) That is, the DM(Bregman) rule can be represented by a decision rule minimizing expected divergence from the true world, which we will call EUM(SP). In the next section, we will explain the reason for calling it that. Although our proof runs along similar lines to the proofs of Banerjee et al. (2005, Theorem 1) and Adamcik (2014, Theorem 2), subtle adjustments are necessary for our belief binarization problem. First, our refined Bregman divergence is defined not only in  $\Delta^W$  but also in  $\Delta^{\mathcal{F}}$ . Second, the refined Bregman divergence is defined neither in  $\mathbb{R}^m$  nor in an open convex subset, but in a closed convex subset  $\Delta^M$ . Third, we allow infinity as a value of divergence.

Throughout, we denote the expectation of  $g$  with respect to a probability distribution  $P \in \mathbb{P}(W)$  by  $\mathbb{E}_{w \sim P}[g(w)]$ , where  $g : W \rightarrow \mathbb{R} \cup \{\infty\}$  or  $g : W \rightarrow \mathbb{R} \cup \{-\infty\}$ . Note that  $\mathbb{E}_{w \sim P}[g(w)] = \sum_{w \in W} P(w)g(w)$ .

**Theorem 2** (Representation of DM(Bregman) by EUM(SP)). *Let  $D$  be a Bregman divergence in  $\Delta^M$ . Then, for all  $p, q \in \Delta^M$  and any probability function  $P \in \mathbb{P}(W)$  represented by  $p$ ,*

$$D(p, q) = \mathbb{E}_{w \sim P}[D(v_w, \cdot)] - \mathbb{E}_{w \sim P}[D(v_w, p)],$$

and thus  $\text{argmin}_b D(p, b) = \text{argmin}_b \mathbb{E}_{w \sim P}[D(v_w, b)]$ .

This theorem states that minimizing distance from  $P$  is the same as minimizing expected distance from the true world according to  $P$  if the distance is given by a Bregman divergence. It is worth noting that we could not prove this theorem if the definition of Bregman divergence guaranteed finiteness only in  $\text{ri}(\Delta^M)$ . Suppose that we use a definition of Bregman divergence that might yield the cases in figure 6, e.g., the definition with (\*) (p. 14) instead of (i) and (ii) in Definition 5. The left-hand

side  $D(p, q)$  is finite for  $q \in \text{ri}(\Delta^M)$ . However,  $D(v_w, p)$  might not be finite even though  $w \in \text{Supp}(P)$ , and thus  $\mathbb{E}_{w \sim P}[D(v_w, p)]$  might not be finite.<sup>9</sup>

The distance between  $v_w$  and  $b$  can be thought of as a utility in epistemic decision theory—an epistemic disvalue. Thus, a DM(Bregman) rule can be seen as a decision rule *maximizing expected utility*. This inspires us to define a new rule that applies epistemic decision theory directly to binarization problems.

## 5. Expected utility maximizing binarization rules

### 5.1. EUM rules with strictly proper scores

Now let us give a formal definition of an *expected utility maximization rule (EUM rule)*.

**Definition 6** (Expected utility maximization rule). *A BR  $G$  is an expected utility maximization rule in  $\Delta^M$  iff there is a utility function  $u : W \times U^M \rightarrow \mathbb{R} \cup \{-\infty\}$  satisfying  $G(P) = \text{argmax}_b \mathbb{E}_{w \sim P}[u(w, b)]$  for all  $P \in \mathbb{P}(W)$ .*

In this paper, we restrict our focus to already well-developed epistemic utility functions, namely proper scores (in Wang (n.d.), we have considered EUM rules with more general utility functions and their properties). This will be useful for investigating the relation between EUM rules and DM(Bregman) rules introduced in the previous section. Put differently, we shall consider the case where  $u(w, b) := -I(w, b)$  for some continuous strictly proper score  $I : W \times \Delta^M \rightarrow [0, \infty]$ , to be defined below.

#### Continuous strictly proper score

Now we define a continuous strictly proper score in our setting where probability distributions are represented in  $\Delta^M$ . We include infinity as a value of scores on some boundary region, and when we talk about continuity of a score including infinity, we will also regulate the region where it should be finite and continuous.

**Definition 7** (Continuous strictly proper score). *Let  $I$  be a function  $I : W \times \Delta^M \rightarrow [0, \infty]$ .*

- (i)  *$I$  is continuous iff, for all  $w \in W$  and  $q \in \Delta^M$ ,*
  - (a) *if  $q \in \mathbb{F}_{v_w}$  then  $I(w, q)$  is finite and continuous in  $q$ ;*
  - (b) *otherwise,  $I$  is extended to a continuous function that might take infinity as a value, meaning that  $(w, q) = \lim_{x \rightarrow q : x \in \mathbb{F}_{v_w}} I(w, x)$ , which exists, infinity being allowed as limits.*
- (ii)  *$I$  is called a strictly proper (SP) score iff, for all  $P \in \mathbb{P}(W)$ ,*

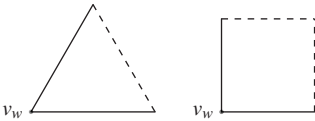
$$\text{argmin}_{q \in \Delta^M} \mathbb{E}_{w \sim P}[I(w, q)] = \{p\},$$

where  $p$  is the representation point of  $P$  in  $\Delta^M$ .

According to our definition above, continuous scores  $I(w, \cdot)$  are finite and continuous not only in  $\text{ri}(\Delta^M)$  but also in  $\mathbb{F}_{v_w}$  (see Figure 7). Notice that  $q \in \mathbb{F}_{v_w}$ ,

<sup>9</sup> We can easily check: (i) if we use (\*) and assume a regularity condition that  $D(v_w, p)$  is finite for all  $w \in \text{MSupp}(p)$  and all  $p \in \Delta^M$ , then we have Theorem 2, and (ii) from the theorem and the regularity condition it follows that  $D(p, q)$  satisfying (\*) is finite for all  $p, q \in \Delta^M$  such that  $p \in \mathbb{F}_q$ .





**Figure 7.** The dashed lines including the end points in each polytope represented where a continuous score  $I(w, \cdot)$  might be infinite. Their complement is  $\mathbb{F}_{v_w}$  where  $I(w, \cdot)$  is finite and continuous.

i.e.,  $w \in \text{MSupp}(q)$ , says that a probability function represented by  $q$  assigns to  $w$  a positive value. This means that  $w$  is probabilistically possible from the point of view of a probability function represented by  $q$ . In this case, we demand that  $I(w, q)$  should not receive infinite score and be continuous in  $q$ . Note that if  $I$  is strictly proper, then we can derive that  $I(w, \cdot)$  is finite in  $\mathbb{F}_{v_w}$ , as shown in the following lemma.

**Lemma 2.** *Let  $I : W \times \Delta^M \rightarrow [0, \infty]$  be a strictly proper score. Then  $I(w, \cdot)$  is finite in  $\mathbb{F}_{v_w}$ .*

It is also worth noting that in  $\Delta^W$  the condition of  $I(w, \cdot)$  being finite in  $\mathbb{F}_{v_w}$  is the same as the notion of *regular* score in Gneiting and Raftery (2007).

*Invariant expectation under the same input representation*

Notice that an expectation value in EUM rules depends not only on the point  $p$  in  $\Delta^M$  but also on the probability distribution  $P$ , in contrast to a divergence in DM rules. Thus, to see the connection between EUM rules and DM rules, we need the following requirement, which is relevant in  $\Delta^{\mathcal{F}}$ , where a point  $p$  might represent several probability distributions.

**Definition 8** (Invariant expectation under the same input-representation). *A function  $I : W \times \Delta^M \rightarrow [0, \infty]$  has an invariant expectation under the same input representation (IER) iff, for all  $P, P' \in \mathbb{P}(W)$  with the same representation in  $\Delta^M$ , i.e.,  $p = p' (\in \Delta^M)$ , we have  $\mathbb{E}_{w \sim P}[I(w, q)] = \mathbb{E}_{w \sim P'}[I(w, q)]$  for all  $q \in \Delta^M$ .*

IER has a close relationship with IIR in Definition 3. If a BR  $G$  is an EUM rule with  $I$  that satisfies IER, then  $G$  is invariant under the same representation (IIR). Although IER may seem a strong condition, it is actually a mild restriction because a large class of scores obey IER. Every scoring function defined in  $\Delta^W$  satisfies IER. We can generalize this to the case in  $\Delta^{\mathcal{F}}$  as follows.

**Lemma 3.** *A function  $I : W \times \Delta^{\mathcal{F}} \rightarrow [0, \infty]$  satisfies IER if  $I$  is a partition-wise score, i.e., there is a partition of  $W$ , say  $W = A_1 \cup \dots \cup A_k$ , such that (i)  $A_1, \dots, A_k \in \mathcal{F}$  and (ii) for all  $i \leq k$  we have, for all  $w, w' \in A_i$  and  $q \in \Delta^{\mathcal{F}}$ ,  $I(w, q) = I(w', q)$ .*

Note that if  $\mathcal{F}$  includes every singleton set  $\{w\}$  of a world, for example,  $\mathcal{P}(W)$ , then every  $I$  is a partition-wise score, and thus it satisfies IER. There are also another ways to satisfy IER. Any additive scores defined in  $\Delta^{\mathcal{F}}$  also enjoy IER, as the following lemma shows. In addition, it shows that strict propriety of an additive score follows from *event-wise strict propriety*.

**Lemma 4.** Let  $I : W \times \Delta^{\mathcal{F}} \rightarrow [0, \infty]$  be additive, i.e., for all  $w \in W$  and  $p \in \Delta^{\mathcal{F}}$ ,

$$I(w, p) = \sum_{A \in \mathcal{F}} I_A((v_w)_A, p_A)$$

where  $I_A : \{0, 1\} \times [0, 1] \rightarrow [0, \infty]$  for all  $A \in \mathcal{F}$ .

- (i)  $I$  satisfies IER.
- (ii) If  $I$  is event-wise strictly proper (E-SP), i.e.,

$$\operatorname{argmin}_{q_A \in [0, 1]} (p_A I_A(1, q_A) + (1 - p_A) I_A(0, q_A)) = \{p_A\}$$

for all  $A \in \mathcal{F}$  and  $p_A \in [0, 1]$ , then  $I$  is strictly proper.

### 5.2. Representation of EUM(SP) by DM(Bregman)

An EUM(SP) is defined as an EUM rule with a continuous strict proper score satisfying IER. Now, let us show how EUM(SP) is related to DM(Bregman). The following theorem shows that an EUM rule with a score  $I$  can be represented by a DM(Bregman) rule when  $I$  is a continuous strictly proper score with IER.

**Theorem 3** (Representation of EUM(SP) by DM(Bregman)). Let  $I : W \times \Delta^M \rightarrow [0, \infty]$  be a continuous strictly proper score with IER. Then there is a Bregman divergence  $D$  in  $\Delta^M$  such that, for all  $p, q \in \Delta^M$  and any probability function  $P \in \mathbb{P}(W)$  represented by  $p$ ,

$$D(p, q) = \mathbb{E}_{w \sim p}[I(w, q)] - \mathbb{E}_{w \sim p}[I(w, p)],$$

and thus  $\operatorname{argmin}_b \mathbb{E}_{w \sim p}[I(w, b)] = \operatorname{argmin}_b D(p, b)$ .

This states that an expected SP score with IER is associated with a Bregman divergence, and minimizing them yields the same result. The next corollary follows from the above theorem.

#### Corollary 1.

- (i) Let  $I$  be a continuous SP score in  $\Delta^W$ . Then  $D(p, q) := \mathbb{E}_{w \sim p}[I(w, q)] - \mathbb{E}_{w \sim p}[I(w, p)]$  is a Bregman divergence in  $\Delta^W$ .
- (ii) Let  $I$  be a continuous additive E-SP score in  $\Delta^{\mathcal{F}}$ . Then  $D(p, q) := \mathbb{E}_{w \sim p}[I(w, q)] - \mathbb{E}_{w \sim p}[I(w, p)]$  is an additive Bregman divergence in  $\Delta^{\mathcal{F}}$ .

We now draw a comparison between our findings and analogous theorems presented in other works. Gneiting and Raftery (2007) and Banerjee et al. (2009) showed similar results to Corollary 1(i). Gneiting and Raftery (2007) established, in  $\Delta^W$ , the relationship between regular proper scores and Bregman divergences. However, scores and divergences are not necessarily assumed to be continuous. Our proof, however, achieves more by deducing the continuity of our refined Bregman divergence through the continuity of scores. The theorem in Banerjee et al. (2009) is similar to the Corollary 1(i), albeit their Bregman divergences are defined on  $\mathbb{R}^m$  rather than  $\Delta^W$ , and they exclude infinity. In contrast, our proof addresses how to

handle infinity. The relation between additive continuous SP scores and additive Bregman divergences in Predd et al. (2009) is similar to Corollary 1(ii). Since they are dealing with non-probabilistic credences as well, their result is stronger than ours in the sense that Bregman divergences are defined on  $[0, 1]^{\mathcal{F}}$  instead of  $\Delta^{\mathcal{F}}$ . However, their findings hinge on the additivity assumption, and in this sense our result is stronger. Compared to the above literature, Theorem 5.2 is more comprehensive in the sense that it provides proofs for the cases in  $\Delta^W$  and in  $\Delta^{\mathcal{F}}$  at the same time. On top of that, Theorem 3 gives us the way to deal with continuous scores in  $\Delta^W$  and non-additive scores in  $\Delta^{\mathcal{F}}$ , at the cost of the assumption of IER.

It is worth asking how our more complicated definition of Bregman divergence plays out in the proof of the theorem. Recall that we could not have had Theorem 2 if we had used the alternative definition (\*) (being finite and continuous in  $\text{ri}(\Delta^M)$ ) instead of parts (i) and (ii) in Definition 5. In contrast, we could have had the same form of Theorem 3 with the alternative notion of Bregman divergence rather than ours. But Theorem 3 tells more with our definition because we demanded more for a divergence to be our Bregman divergence and thus we proved more.

From Theorems 3 and 2, we have the following claims, which might be viewed as involving a converse of both theorems in certain conditions.

**Corollary 2.**

(i) Let  $I : W \times \Delta^M \rightarrow [0, \infty]$  satisfying IER.  $I$  is continuous SP iff

$$D_I(p, q) := \mathbb{E}_{w \sim p}[I(w, q)] - \mathbb{E}_{w \sim p}[I(w, p)]$$

is a Bregman divergence.

(ii) Let  $D : \Delta^M \times \Delta^M \rightarrow [0, \infty]$  be a divergence, and suppose that  $I_D(w, q) := D(v_w, q)$  satisfies IER.  $D$  is a Bregman divergence iff

$$D(p, q) = \mathbb{E}_{w \sim p}[D(v_w, q)] - \mathbb{E}_{w \sim p}[D(v_w, p)]$$

and  $I_D(w, q)$  is continuous in  $q$ .

To summarize this section and the previous one, we proved, with our refined definitions, that the DM(Bregman) rules and the EUM(SP) rules have the same extension under certain conditions (IER): (i) a strictly proper score  $I$  of an EUM rule satisfying IER can be extended to a Bregman divergence  $D_I$  such that  $D_I(p, q) = \mathbb{E}_{w \sim p}[I(w, q)] - \mathbb{E}_{w \sim p}[I(w, p)]$ , and the DM rule with  $D_I$  generates the same results with the EUM rule; (ii) a Bregman divergence  $D$  of a DM rule can be restricted to a strictly proper score  $I_D$  such that  $I_D(w, q) = D(v_w, q)$ , and the EUM rule with  $I_D$  generates the same results with the DM rule.

**6. Conclusion**

Numerous questions still remain unanswered. Although our belief binarization methods can be applied to both prior and posterior, we do not presuppose or advocate for any dynamic norms for credence, such as plan conditionalization as in Greaves and Wallace (2006) or Laplacian imaging as in Leitgeb and Pettigrew (2010), and dynamic norms for the belief-credence connection, such as the commutativity norm shown as in Lin and Kelly (2012a) or the compatibility norm concerning AGM belief revision as

in Leitgeb (2017). Based on our study, an intriguing question arises: Can we identify or characterize EUM rules that track Bayesian (plan) conditionalization by employing certain suitable rational belief revision methods? This question has to be left for a different paper.

The DM(Bregman) rules include an infinite number of distance measures. By incorporating additional rationality norms, we can identify a more appealing subset of DM(Bregman). One such property would be that the probability of every believed proposition generated by some belief binarization rule ought to be bigger than half. In Wang and Kim (*n.d.*), we showcase the fulfillment of this criterion by DM(SE). Exploring broader discoveries in this vein would indeed be intriguing.

**Supplementary material.** For supplementary material accompanying this paper visit <https://doi.org/10.1017/psa.2024.9>

**Acknowledgements.** We sincerely thank Hannes Leitgeb and Christian List for invaluable feedback on a previous version. Our paper was presented at the Formal Epistemology Workshop 2022, where commentator Hanti Lin and the audience provided insightful feedback, enhancing the focus and clarity of our work. Special gratitude to two anonymous reviewers for their detailed discussion points. The first author acknowledges support from Studienstiftung des Deutschen Volkes.

## References

- Adamcik, Martin. 2014. "Collective reasoning under uncertainty and inconsistency." PhD diss., University of Manchester.
- Banerjee, Arindam, Xin Guo, and Hui Wang. 2005. "On the optimality of conditional expectation as a Bregman predictor." *IEEE Transactions of Information Theory* 51 (7):2664–9. <https://doi.org/10.1109/TIT.2005.850145>.
- Chandler, Jake. 2013. "Acceptance, aggregation and scoring rules." *Erkenntnis* 78:201–17. <https://doi.org/10.1007/s10670-012-9375-6>.
- Dietrich, Franz, and Christian List. 2017a. "Probabilistic opinion pooling generalized. Part one: General agendas." *Social Choice and Welfare* 48:747–86. <https://doi.org/10.1007/s00355-017-1034-z>.
- Dietrich, Franz, and Christian List. 2017b. "Probabilistic opinion pooling generalized. Part two: The premise-based approach." *Social Choice and Welfare* 48:787–814. <https://doi.org/10.1007/s00355-017-1035-y>.
- Dietrich, Franz, and Christian List. 2021. "The relation between degrees of belief and binary beliefs: A general impossibility theorem." In *Lotteries, Knowledge, and Rational Belief: Essays on the Lottery Paradox*, edited by Igor Douven, 223–54. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108379755.012>.
- Dorst, Kevin. 2019. "Lockeans maximize expected accuracy." *Mind* 128 (509):175–211. <https://doi.org/10.1093/mind/fzx028>.
- Gneiting, Tilmann, and Adrian E. Raftery. 2007. "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102 (477):359–78. <https://doi.org/10.1198/016214506000001437>.
- Goodman, Jeremy, and Bernhard Salow. 2023. "Epistemology normalized." *Philosophical Review* 132 (1): 89–145. <https://doi.org/10.1215/00318108-10123787>.
- Greaves, Hilary, and David Wallace. 2006. "Justifying conditionalization: Conditionalization maximizes expected utility theory." *Mind* 115 (459):607–32. <https://doi.org/10.1093/mind/fzl607>.
- Hempel, Carl G. 1960. "Inductive inconsistencies." *Synthese* 23 (4):439–69. <https://doi.org/10.1007/BF00485428>.
- Joyce, James M. 1998. "A nonpragmatic vindication of probabilism." *Philosophy of Science* 65 (4):575–603. <https://doi.org/10.1086/392661>.
- Leitgeb, Hannes. 2014. "Belief as a simplification of probability, and what this entails." In *Johan van Benthem on Logic and Information Dynamics*, edited by Alexandru Baltag and Sonja Smets, 405–17. New York: Springer. [https://doi.org/10.1007/978-3-319-06025-5\\_14](https://doi.org/10.1007/978-3-319-06025-5_14).

- Leitgeb, Hannes. 2017. *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198732631.001.0001>.
- Leitgeb, Hannes, and Richard Pettigrew. 2010. "An objective justification of Bayesianism II: The consequences of minimizing inaccuracy." *Philosophy of Science* 77 (2):236–72. <https://doi.org/10.1086/651318>.
- Levi, Isaac. 1967. *Gambling with the Truth: An Essay on Induction and the Aims of Science*. Cambridge, MA: MIT Press.
- Lin, Hanti, and Kevin T. Kelly. 2012a. "Propositional reasoning that tracks probabilistic reasoning." *Journal of Philosophical Logic* 41 (6):957–81. <https://doi.org/10.1007/s10992-012-9237-3>.
- Lin, Hanti, and Kevin T. Kelly. 2012b. "A geo-logical solution to the lottery paradox." *Synthese* 186 (2): 531–75. <https://doi.org/10.1007/s11229-011-9998-1>.
- Oddie, Graham. 1997. "Conditionalization, cogency, and cognitive value." *British Journal of the Philosophy of Science* 48 (4):533–41. <https://doi.org/10.1093/bjps/48.4.533>.
- Pettigrew, Richard. 2015. "Accuracy and the credence–belief connection." *Philosophers' Imprint* 15 (26):1–20.
- Predd, Joel B., Robert Seiringer, Elliott H. Lieb, Daniel N. Osherson, H. Vincent Poor, and Sanjeev R. Kulkarni. 2009. "Probabilistic coherence and proper scoring rules." *IEEE Transactions of Information Theory* 55 (10):4786–92. <https://doi.org/10.1109/TIT.2009.2027573>.
- Rockafellar, Ralph T. 1970. *Convex Analysis*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400873173>.
- Wang, Minkyung. (n.d.) "Credence and belief: Epistemic decision theory revisited." Unpublished manuscript.
- Wang, Minkyung, and Chisu Kim (n.d.) "Belief as a Natural Concept." Unpublished manuscript.