DOMINIK KLEIN
ONDREJ MAJER
& SOROUSH RAFIEE RAD

NON-CLASSICAL PROBABILITIES FOR DECISION MAKING IN SITUATIONS OF UNCERTAINTY

INTRODUCTION

Classical probability gives a quantitative representation of uncertainty over a given space of available events. It assigns to each event a value in the interval $[0,1]$ interpreted as some agent's subjective credence therein or, alternatively, its likelihood of occurrence. The space of events, here, is assumed to have an internal structure, where some events are simple and others compound. Overall, the different events can be arranged into a Boolean algebra, a set-theoretical structure with union, intersection and complementation.

The probabilities that can be assigned to different events stand in various relations to each other. It seems incompatible, for instance, to assign a larger probability (or credence) to some specific event B than to some more general

DOMINIK KLEIN, PhD, Utrecht University, Department of Philosophy and Religious Studies; e-mail: d.klein@uu.nl; ORCID: https://orcid.org/0000-0002-7743-8399.

ONDREJ MAJER, PhD, Institute of Philosophy of the Czech Academy of Sciences; e-mail: majer@flu.cas.cz; ORCID: https://orcid.org/0000-0002-7243-1622. The research of Ondrej Majer was supported by the grant GA16-15621S.

SOROUSH RAFIEE RAD, PhD, Bayreuth University, Department of Philosophy; e-mail: Soroush. R.Rad@gmail.com; ORCID: https://orcid.org/0000-0001-5338-902X.

A that encompasses B. The standard assumptions on probabilities are captured in Kolmogorov's axioms (Kolmogorov). These are stated for *probability functions*, a normalized additive measure defined on a Boolean algebra F of subsets[1] of some set W, representing some agent's subjective representation of the world. That is, a probability measure is a function $\mu : F \to \mathbb{R}$ satisfying the three Kolmogorov's axioms that for every $A, B \in F$

1. $\mu(A) \geq 0$ (non-negativity)
2. $\mu(W) = 1$ (normalization)
3. $\mu(A \cup B) = \mu(A) + \mu(B)$, , for disjoint A, B (additivity)

Similarly, probability measures can be defined syntactically, on a set of formulas of some (classical) propositional language L. In this perspective, a probability measure is a function $\mu : \mathcal{L} \to \mathbb{R}$ and the above axioms can be rephrased as (i) $\mu(\varphi) \geq 0$ (non-negativity), (ii) $\mu(\top) = 1$ (normalization), and (iii) $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$, for inconsistent $\phi \wedge \psi$ (additivity), which are to be accompanied by an additional axiom (iv) $\mu(\varphi) = \mu(\psi)$ whenever $\vdash \varphi \leftrightarrow \psi$.

Defined as such the classical probability has strong consistency and completeness assumptions built in. Using axioms above, it is not difficult to show for any sentence $\varphi \in \mathcal{L}$ that, $P(\varphi \wedge \neg\varphi) = 0$, $P(\varphi \vee \neg\varphi) = 1$ and $P(\varphi) + P(\neg\varphi) = 1$. These mathematically desirable properties, however, come at a cost: by incorporating various normative coherence assumptions, classical probability becomes a less than optimal tool for representing quantitative uncertainty in non-idealised settings where evidence may be incomplete or partially faulty. When information is all too scarce, for instance, the resulting probabilities for the truth and falsity of some φ may add up to less than one. Likewise, an occasional piece of faulty input may lead to contradictory beliefs and thus, by composition, to $P(\varphi \wedge \neg\varphi) > 0$.

The issue goes deeper than the definition of probability however. The underlying logic for classical probability theory is classical propositional logic which already assumes these consistency and completeness properties. Indeed, as is well-known, any inference in classical logic is trivialized when dealing with inconsistent premises. This stands in stark contrast to the fact that we can often draw at least *some* valid inferences from a body of partially contradictory information.

The adequacy of classical logic for modelling and analyzing information dynamics has been questioned for quite some time now, not only on the grounds of its inadequacy for dealing with inconsistencies. To date, a broad

---

[1] For the sake of simplicity, we assume the state space W to be finite in this article. Without this assumption, the additivity axiom would need to be replaced by σ-additivity.

variety of logical systems have been been put forward that extend or modify classical logic in order to reason with uncertain degrees of truth (PARIS 1994; FAGIN, HALPERN, and MEGIDDO 1990), incorporate epistemic notions that bear relevance to reasoning about others (VAN BENTHEM 2011; VAN DITMARSCH, VAN DER HOEK, and KOOI 2008; VAN BENTHEM, VAN EIJCK, and KOOI 2006), represent the dynamics of information (ALCHOURRÓN, GÄRDENFORS, and MAKINSON 1985; BALTAG, MOSS, and SOLECKI 1998; BALTAG, RENNE, and SMETS 2014; DARWICHE, PEARL 1997), or track various aspects of reasoning processes (BATENS 2001; ANDERSON, BELNAP 1975), to name but a few.

Most relevant to our discussion here, are extensions that provide a non-trivial treatment of inconsistencies (PRIEST, 2006, 2007; BELNAP 2019; DUNN 1976; JAŚKOWSKI 1989; DA COSTA 1974; DA COSTA, and SUBRAHMANIAN 1989; RESCHER and MANOR 1970; BATENS 2001). By now, the literature on para-consistent logical systems is extensive and contains many different conceptual and formal approaches to inconsistencies. Some of these treat inconsistencies as inherent in reality, most notable in the work of Graham Priest (PRIEST 1979; 2006; 2007), while others view it as a result of information shortcomings of different kinds, see for example (BELNAP 1977; 2019; DUNN 1976; JAŚKOWSKI 1969; DA COSTA 1974; RESCHER and MANOR 1970; BATENS 2001; ANDERSON and BELNAP 1975). In this paper we will focus primarily on Belnap and Dunn's view and their resulting logic, *Belnap-Dunn logic* or *First Degree Entailment*. By permitting information to be inconsistent or incomplete, this logic track how and which inferences can be drawn from an imperfect body of evidence.

Informational imperfections are, of course, not restricted to the case of categorial *true-false* information and deductive reasoning. Rather, similar issues of inconsistent, conflicting or incomplete inputs may also emerge when dealing with probabilistic evidence. Information that stems from different sources might, for instance, indicate probabilities for φ and $\neg\varphi$ that add up to more, or less than 1. In the same way in which probability theory can be seen as an extension of classical logic to uncertain information, this calls for an extension of paraconsistent logic to a weaker than standard probability theory.

In this paper we focus on a probabilistic extensions defined over Belnap-Dunn logic. This paraconsistent logic is particularly suitable for analyzing less-than-ideal informational situations. Not only does it allow for handling of inconsistencies but it also relaxes the assumption of information completeness. More precisely in Belnap-Dunn logic $\phi \vee \neg\phi$ is no longer a logical validity. This allows for modelling not only the epistemic state of an

agent with contradictory evidence for both ϕ and $\neg\phi$, but also that of an agent with evidence neither for ϕ nor for $\neg\phi$.

There has been some recent works on non-standard probability that rest on Belnap-Dunn logic implicitly or explicitly (cf. CHILDERS, MAJER, and MILNE 2019; DUNN 2010; DUNN and KIEFER 2019; KLEIN, MAJER, and RAFIEE RAD 2020). Childers, Majer, and Milne (2019), for instance, obtain a non-standard probability theory by relaxing some of Kolmogorov's axioms. As we shall see in detail shortly, their approach weakens both the axioms for unity ($P(\top) = 1$) and (σ)-additivity. A different proposal has been put forward by Dunn and Kiefer (2019). They define non-standard probabilities as a four valued assignments that associates to each proposition ϕ a normalized vector $(b, d, u, c) \in [0,1]^4$, whose values correspond to "pure belief" (ϕ is true while $\neg\phi$ is not), "pure disbelief" ($\neg\phi$ is true and ϕ is not), "uncertainty" (neither ϕ nor $\neg\phi$ are true), and "conflict" (ϕ and $\neg\phi$ are both true). Dunn's proposal is studied in detail and extended in Klein, Majer, and Rafiee Rad (2020) where the setting is enriched with dynamical operations for conditionalization and aggregation to capture a process of probabilistic learning. Moreover, Klein, Majer, and Rafiee Rad (2020) show the non-standard probabilities à la Childers, Majer, and Milne (2019), and the four valued probabilities proposed by Dunn and Kiefer (2019) to be equivalent perspectives on the same, and provide an explicit translation between the two. Using this, we shall focus here on the (single-valued) non-standard probabilities as defined in Childers, Majer, and Milne (2019).

Notably, employing (non-classical) probabilities is not the only approach in the literature for dealing with imperfections in a quantitative representation of uncertainty. The most discussed formal systems in this area are inner and outer measures, lower and upper probabilities and Dempster-Shafer theory. Among these the last is particularly prominent, employing *belief functions* to represent uncertainty within quantitative belief assignments (DEMPSTER 1967; SHAFER 1976). In a certain technical sense it is also supposed to be the most general of the theories we mentioned (see e.g. HALPERN 2017). Dempster-Shafer theory has been developed as framework for a theory of evidence. In this setting, events (or propositions) are assigned values in $[0,1]$ that express how much the event is supported by the available evidence. Within the DS-framework, evidence is assumed to be non-misleading and to follow some classic meta-theory. That is, the empty set of propositions receives a support of 0 while the set of all events is believed to a degree of 1. Evidence, however, may be partial in that the evidence implying some set U

and the evidential support for its complement may only be a small fraction of the overall evidence available to the agent. In this case, the corresponding belief function will satisfy $\mathrm{bel}(U) + \mathrm{bel}(\neg U) < 1$. Dempster-Shafer theory has been extensively analyzed as one of the main proposals for modelling epistemic uncertainty. See for instance (HALPERN 2017) for an extensive introduction.

The goal of this paper is to provide a comparison between Dempster-Shafer's theory as the most general and most prominent framework for dealing with imperfect information and non-standard probabilities[2] in terms of their perspectives on informational imperfections. Another reason that makes this comparison relevant is the strong connection between information imperfections and evidence. Incomplete and inconsistent evidence gathered from conflicting sources are some of the main contributers to information imperfections that agents should deal with when forming belief. Hence, it seems a natural task to compare setting that allow for agents to form beliefs based on imperfect information with the main theory of evidence available in the literature.

In this comparison, we are in particular interested in two questions. The first is to compare the two approaches in terms of their generality. More specifically we ask whether one is more general than the other, whether there is a (partial) correspondence between the two or whether (and under what conditions) either of these approaches can be seen as a special case of the other. To subsume, we are interested whether or not (and under what conditions) a non-standard probability function can be interpreted as a Dempster-Shafer belief function and vice versa. Second, we wish to extend our comparison of both approaches to dynamics and their respective mechanisms for learning and information change. For either of the two approaches dynamic rules have been proposed that are claimed to generalize classic probability's Bayes update (cf. SHAFER 1976; FAGIN and HALPERN 1990 for Dempster-Shafer theory and JØSANG 1997; KLEIN, MAJER, and RAFIEE RAD 2020 for the non-standard case). We ask how these rules compare to each other and whether they are compatible with any (partial) correspondence to be found in reply to the first question.

The paper is organized as follows. In section 1 we provide brief introductions to Dempster-Shafer theory (section 1.1) and non-standard probability theory (Section 1.2.2). In Section 2 we then analyze the relationship

---

[2] Or, by the translation in KLEIN, MAJER and RAFIEE RAD 2020, the four-valued probabilities.

between the two frameworks. While we show the two frameworks to be orthogonal, strictly speaking, (Section 2.1), we also identify a partial bridge between them in Section 2.2. With the help of this partial bridge, we then compare various dynamic rules for the frameworks that are all intended to generalize Bayes in Section 3 before concluding.

1. PRELIMINARIES

To begin with, we rehearse the basics of Dempster-Shafer theory and non-standard probability theory. Here, we focus on the basics needed for our analysis. For a more complete introduction, we refer the reader to Dempster (1967) and Halpern (2017) for Dempster-Shafer theory and Klein, Majer, and Rafiee Rad (2020) for non-standard probability theory.

1.1 DEMPSTER-SHAFER THEORY

Originally, Dempster-Shafer (DS) theory has been developed as a theory of evidence-based belief. More specifically, the DS belief in some event is derived form the agent's evidential support for various propositions. In this setting, evidential support for some proposition U tracks the amount of evidence for precisely U. That is, the support of U reflects all the evidence that support U, but no proposition stronger than U. The evidential support for some U is, in particular, completely independent of the evidential support for any event V that implies U, i.e. $V \subseteq U$. An agent's belief, on the other hand, builds on logical entailment among various propositions. More concretely, the agents belief in some U is calculated from her evidential support for all propositions V that logically entail U.

Take for example the DS degree of belief about the colour of an observed object. The degree of belief that the object is either blue or green depends on the evidence that support *exactly* the assertion that the object is *either blue or green* together with the evidence that support the object being either only blue or only green. The agents beliefs that the object is blue (resp. green), on the other hand, only depend on the respective evidential supports for the objects being blue (resp. green). Thus, the agent's belief that an object is "either blue or green" can be higher than the sum of her beliefs in the object being "blue" and "green," as the former rests on all the evidence contained in the latter, as well as on additional evidence for the object being "blue or

green." More specifically, as soon as the agent is exposed to some piece of (exact) evidence for the object being "blue or green," her belief in the disjunction will be higher than the sum of beliefs in the disjuncts. Formally, a DS belief function captures the amount of available evidence by means of a mass function m. Belief is then defined with recourse to this mass function.

Definition 1 Let W be a set (of alternatives) and $\mathcal{P}(W)$ the set of events over W.

(i) A Dempster-Shafer model is a tuple (W, m) where $m : \mathcal{P}(W) \to [0,1]$ such that (i) $m(\varnothing) = 0$, and (ii) $\sum_{A \subseteq W} m(A) = 1$.

Intuitively, m is a mass function which assigns to every event (or proposition) $A \subseteq W$ the weight of evidence which supports *exactly* A, i.e A and no subset of A.

(ii) Every DS model induces a *belief function* $\mathrm{bel} : \mathcal{P}(W) \to [0,1]$

$$\mathrm{bel}(A) := \sum_{A' \subseteq A} m(A')$$

The belief function of a DS model represents the belief warranted in the various subsets of W, say A, in light of the available evidence. This belief is equated to the *total evidential support* for this A, i.e. the sum of evidential supports for all subsets of A. In the above definition, a DS belief function is a mapping from the Boolean algebra of subsets[3] of some set W to the interval $[0,1]$ that is characterized by the following axioms:

$$i) \ \mathrm{bel}(\varnothing) = 0 \qquad\qquad ii) \ \mathrm{bel}(W) = 1$$

$$iii) \ \mathrm{bel}(A_1 \cup \ldots \cup A_n) \geq \sum_{I \subseteq \{1 \ldots n\}} (-1)^{|I|+1} \mathrm{bel}(\cap_{i \in I} A_i)$$

Dempster-Shafer theory can be seen as a generalization of Bayesian framework of subjective belief. In fact, a DS-belief function is a probability measure if and only if each piece of evidence supports a singleton $x \in W$, i.e. the underlying mass function satisfies $m(X) = 0$ whenever $X \subseteq W$ is not a singleton. A crucial aspect of such theories, both conceptually and formally, is the concept of conditionalization (HARTMANN and SPRENGER 2019). For its main part, conditionalization extend the scope of the theory from being a static representation of uncertainty to a dynamic one by pro-

---

[3] Generalizations exists that, for instance, work on distributive lattices rather than algebras of sets – see ZHOU 2013.

viding a mechanism for the incorporation of new evidence. This addition is crucial for analyzing information dynamics and its related uncertainties, thereby extending the realm of possible applications of the theory. Often, it is not only the agents' static epistemic states that are of interest, but various of their dynamic properties, such as which information may have lead to an agents' current beliefs, how various actual or potential observations should impact our credences or where beliefs converge to in the long run. All these topics are essentially dynamic, requiring a formal tool for predicting and assessing the effects of informational updates.

We will not delve into a detailed representation of DS theory and its ramifications here. Besides the absolute basics just introduced, we shall restrict ourselves to presenting two notions of conditionalization for DS belief functions that have been proposed in the literature. However, we will postpone this to Section 3, where we compare these two dynamics on DS belief functions with a version of Bayes updating defined on non-standard probabilities.

1.2 NON-STANDARD PROBABILITIES

There is a second, logic based approach for representing quantitive beliefs that supervene on possibly imperfect information. In a nutshell, the idea of these *non-standard probability functions* is as follows. At the outset, the structure of belief under sub-ideal information is discussed when all available information and their effects are extremal, i.e. assume the value 0 or 1. For classic probability theory, the restrictions to such extremal values yields a classic propositional logic. The transition from idealized classic to sub-ideal accounts of input information, will thus yield a weakening of propositional logic. In Section 1.2.1, we introduce a particularly prominent weakening of propositional logic, Belnap-Dunn logic, that has been specifically designed to capture cases of incomplete and inconsistent information. In the section thereafter, 1.2.2, we present recent approaches for generalizing Belnap-Dunn logic to the non-extremal values, i.e. to define real-valued probability functions whose restrictions to their extremal values (i.e. their underlying logic) yields Belnap-Dunn logic.

1.2.1 *Belnap-Dunn logic*

By now, there is a large number of proposals for paraconsistent logics, i.e. logics where local contradictions do not trigger trivialization. Among

these is Belnap-Dunn logic (BD-logic, also referred to as First Degree Entailment) that was specifically designed for dealing with inferences from (possibly) incomplete and inconsistent information, as can be found for instance in databases. This is accomplished by extending the set of truth values. In BD a formula might not only be true (t) or false (f) as in the classical case, but can also be neither (n) or both (b) of these, where b and n correspond to inconsistent and incomplete information respectively.[4]

An alternative to four valued semantics for DB logic consists in a double valuation approach. While employing the classic set of truth values $\{t, f\}$, this formulation builds on the assumption that positive and negative evidence are distinct. That is, the absence of positive evidence for some p is not the same as negative evidence for p (or positive evidence for $\neg p$, if you will). In fact, both are completely independent. That is, there may be gaps, where neither evidence for p nor against p is available, and gluts, where evidence of both is present. Consequently, positive and negative evidence require a separate treatment. This is achieved by duplicating the valuation function. Instead of a single valuation function telling us which atomic propositions are true and which false, we have to work with two valuation functions — the positive one telling us for which atomic propositions we have positive evidence and a negative one corresponding to available evidence against atomic propositions.

Independent of the semantic approach, the language $\mathcal{L}$ of BD logic is generated by $\wedge, \vee, \neg$ from a finite set of atomic letters At. We will denote the set of literals over At by Lit, i.e. $\mathrm{Lit} := \mathrm{At} \cup \{\neg p \mid p \in \mathrm{At}\}$. In this paper, we will focus on a semantic perspective of BD-logic. That is, instead of providing a syntactic entailment relation, we define a set of models and derive semantic entailment relations $\vDash$ over these. As positive and negative evidence for (atomic) propositions are assumed independent, models of BD-logic require two separate valuation functions v^+ and v^- for positive and negative evidence respectively.

Definition 2 A *non-standard model* is a triple $\mathcal{M} = \langle \Sigma, v^+, v^- \rangle$ where Σ is a finite or countably infinite set of states and $v^+, v^- : \Sigma \times \mathrm{At} \to \{0, 1\}$ are the positive and negative valuation function respectively. For $p \in \mathrm{At}$ we let $v^{\pm}(p) = \{s \in \Sigma | v^{\pm}(s, p) = 1\}$.

---

[4] For our purpose, the reading of truth values is not metaphysical as e.g. in dialethism. Rather, they are interpreted with respect to the available information. Or, as Belnap puts it "True" is to be read as "Told true".

With this definition, a state s of a model $\mathcal{M}$ might be non-classic. That is, it can be assigned an inconsistent set of propositions ($s \in v^+(p) \cap v^-(p)$ for some $p \in Prop$), or might be incomplete with respect to other propositions ($s \notin v^+(q) \cup v^-(q)$ for some $q \in Prop$). While positive and negative valuations are defined on atoms only, they can be extended to compound formulas in a way that is borrowed from classical logic – a state supports conjunction positively if it positively supports both conjuncts and supports it negatively if it negatively supports either. Formally, this is made precise by positive and negative support relations $\vDash^+$ and $\vDash^-$.

Definition 3 Let $\mathcal{M} = \langle \Sigma, v^+, v^- \rangle$ be a non-standard model, $s \in \Sigma$ a state and $\varphi, \psi \in \mathcal{L}$ formulas. Then the *semantics* of $\mathcal{L}$ is given by:

$$
\begin{array}{ll}
s \vDash^+ p \text{ iff } s \in v^+(p) & s \vDash^- p \text{ iff } s \in v^-(p) \\
s \vDash^+ \varphi \wedge \psi \text{ iff } s \vDash^+ \varphi \text{ and } \mathcal{M}, s \vDash^+ \psi & s \vDash^- \varphi \wedge \psi \text{ iff } s \vDash^- \varphi \text{ or } s \vDash^- \psi \\
s \vDash^+ \neg\varphi \text{ iff } s \vDash^- \varphi & s \vDash^- \neg\varphi \text{ iff } s \vDash^+ \varphi
\end{array}
$$

Building on this semantics, we obtain the notions of positive and negative extensions. The *positive extension* of a formula $\varphi \in \mathcal{L}$ is the set of states positively supporting it, $\mid \varphi \mid^+_{\mathcal{M}} = \{s \in \Sigma \mid \mathcal{M}, s \vDash^+ \varphi\}$, and its *negative extension* is $\mid \varphi \mid^-_{\mathcal{M}} = \{s \in \Sigma \mid \mathcal{M}, s \vDash^- \varphi\}$. Observe that positive and negative extensions are inter-definable via negation: $\mid \varphi \mid^-_{\mathcal{M}} = \{s \in \Sigma \mid \mathcal{M}, s \vDash^+ \neg\varphi\} = \mid \neg\varphi \mid^+_{\mathcal{M}}$. Thus it is sufficient to work with positive extensions only. Entailment relation between sentences are defined in the usual way: $\phi \vDash^+ \psi$ if and only if for all models $\mathcal{M}$ and states s, if $\mathcal{M}, s \vDash^+ \phi$ then $\mathcal{M}, s \vDash^+ \psi$, likewise for $\phi \vDash^- \psi$.

We end this section by noting a fact that we will need later: Belnap-Dunn logic has no tautologies, but the top-element of its Lindenbaum algebra is $\bigvee_{a \in \text{Lit}} a$, i.e. $\varphi \vDash \bigwedge_{a \in \backslash \text{Lit}} a$ for all φ.

1.2.2 *Probabilities over Belnap-Dunn logic*

Now, we are in a position to introduce probabilities over Belnap-Dunn logic. To do so, we will assume the non-standard models defined above to be equipped with a (classical) probability measure μ defined on the powerset of the set of states Σ. Just as in the classical case, the non-standard probability of a formula φ will be defined as the probability given by μ to states that positively support it. Notice that these are the states in which φ is *at least* true, since a state might simultaneously also positively support $\neg\varphi$.

Definition 4 A *probabilistic model* is a tuple $\mathcal{M} = \langle \Sigma, v^+, v^-, \mu \rangle$ where $\langle \Sigma, v^+, v^- \rangle$ is a non-standard model and μ is a (classical) probability measure on the full subset algebra of Σ. The *non-standard probability* function $P_{\mathcal{M}}^+ : \mathcal{L} \to \mathbb{R}$ induced by μ is:

$$P_{\mathcal{M}}^+(\varphi) = \mu(\mid \varphi \mid_{\mathcal{M}}^+)$$

Analogously, we could define the negative probability of some φ. However, due to the interdefinability of positive and negative support, $\mid \varphi \mid_{\mathcal{M}}^- = \mid \neg\varphi \mid_{\mathcal{M}}^+$, this is not necessary. By the same reasoning, we will omit the subscript $^+$ when talking about probability. It can be checked that for every probabilistic model $\mathcal{M} = \langle \Sigma, v^+, v^-, \mu \rangle$, the corresponding non-standard probability function $P_{\mathcal{M}}$ satisfies the following axioms.

(A1)	$0 \leq P(\varphi) \leq 1$	(normalization)
(A2)	if $\varphi \vDash_L \psi$ then $P(\varphi) \leq P(\psi)$	(monotonicity)
(A3)	$P(\varphi \wedge \psi) + P(\varphi \vee \psi) = P(\varphi) + P(\psi)$	(inclusion/exclusion)

where $\vDash_L$ in (A2) is the entailment relation of Belnap-Dunn logic defined above. In fact, these axioms are sound and complete with respect to the class of probabilistic models defined above (KLEIN, MAJER, and RAFIEE RAD 2020). Thus, we call a function $p : \mathcal{L} \to \mathbb{R}$ satisfying (A1)-(A3) a *non-standard probability assignment*. Notably, the axioms for non-standard probability are strictly weaker than Kolmogorov's. Axioms (A1)-(A3) can be derived from the Kolmogorov axioms. This follows from the fact that first degree entailment is a sub-relation of classical entailment (KLEIN, MAJER, and RAFIEE RAD 2020). In the other direction, however, we can only get the non-negativity axiom ($P(\varphi) \geq 0$ for all φ) from (A1). Neither the unit axiom ($P(\top) = 1$) nor the (σ)-additivity axioms can be derived from (A1)-(A3). To see this notice that assigning probability .5 to *all* formula satisfies axioms (A1)-(A3).

Notably, the probabilities of φ and $\neg\varphi$ need not sum up to 1 in this setup. The only constrain on the relation between the probability of a formula and its negation is given by the inclusion-exclusion rule(A3):

$$P(\varphi \vee \neg\varphi) + P(\varphi \wedge \neg\varphi) = P(\varphi) + P(\neg\varphi)$$

which allows for both (probabilistic) gaps ($P(\varphi \vee \neg\varphi) < 1$) and gluts ($P(\varphi \wedge \neg\varphi) > 0$).

To end this section, we define a special class of probabilistic models called canonical models. We will need these later, for producing a partial bridge result between DS theory and non-standard probability.[5]

Definition 5 i) A canonical probabilistic model is a probabilistic model $\mathcal{M} = \langle \mathcal{P}(Lit), v^+, v^- \mu \rangle$ such that $v^+(s,p) = 1$ iff $p \in s$ and $v^-(s,p) = 1$ iff $\neg p \in s$.

ii) $\mathbb{M}^{\mathbf{can}}$ is the set of canonical probabilistic models.

One reason for calling these models *canonical* is that there is a tight relationship between canonical probabilistic models and non-standard probability assignments (and thus also the non-standard probability functions derived from probabilistic models).

Theorem 1 (Theorem 4 of of KLEIN, MAJER, and RAFIEE RAD 2020) *The function f from $\mathbb{M}^{\mathbf{can}}$ to the set of all non-standard probability assignments on $\mathcal{L}$ mapping a canonical probabilistic model $\mathcal{M}$ to $P_{\mathcal{M}}$ is a bijection.*

In other words. We can represent each non-standard probability assignment *uniquely* with a corresponding canonical probabilistic model. We will make heavy use of this later.

2. DEMPSTER-SHAFER THEORY AND NON-STANDARD PROBABILITIES

In this section, we initiate a comparison between Dempster-Shafer Theory and non-standard belief functions. As it will turn out, they model *different* aspects of evidential incompleteness or incoherence.

To begin with, we note that Dempster Shafer Theory allows for belief gaps, i.e. cases where $bel(A) + bel(\neg A) < 1$. However, it does not allow for gluts, i.e. for cases where $bel(A \wedge \neg A) > 0$ or $bel(A) + bel(\neg A) > 1$. To see the latter, note that for all $B \subseteq A$, $B \subseteq W$ and similarly for all $B \subseteq \neg A$, $B \subseteq W$. We then obtain

$$1 = \mathrm{bel}(W) = \sum_{B \subseteq W} m(B) \geq \sum_{B \subseteq A} m(B) + \sum_{B' \subseteq \neg A} m(B') \geq \mathrm{bel}(A) + \mathrm{bel}(\neg A)$$

This implies that DS theory can handle situations of incomplete information but is unsuitable for situations of inconsistent evidence. In this sense

---

[5] This definition is a slight simplification from the one presented in (KLEIN, MAJER, and RAFIEE RAD 2020).

non-standard probabilities are more general from the outset. Moreover, non-standard probabilities also allow for more general types of gaps than DS theory does. In the latter, agents cannot be less than fully certain about classical tautologies, that is $\mathrm{bel}(\varphi) = 1$ whenever φ is a tautology of classical logic. Non-standard probabilities, in contrast, allow for such uncertainty. In the extreme case, the non-standard belief function of maximal uncertainty sets $P(\varphi) = 0$ for *every* $\varphi \in \mathcal{L}$. An immediate consequence of this is that Dempster-Shafer belief functions are not more general than non-standard probabilities.

However, it is also not true that non-standard probabilities are strictly more general than DS belief functions. Rather, Dempster-Shafter theory allows for gluts of the form $\mathrm{bel}(A) + \mathrm{bel}(B) < \mathrm{bel}(A \vee B)$. Intuitively, these are cases where the agent has a reliable information that $A \vee B$, while only having weak information for either disjunct. In our example above, this reflects a case where some of the agent's available evidence is that the object observed was 'blue or green", without specifying with of these two was the case. Notably, such a situation cannot occur within non-standard belief functions, as the inequality $P(A) + P(B) < P(A \vee B)$ violates the import-export rule (A3). Hence, non-standard belief functions are not more general than DS theory either.

2.1 AN ORTHOGONALITY RESULT

In a certain sense, this discrepancy mirrors different motivations under-lying DS theory and non-standard belief functions. In DS theory, there are no morphological limitations to the available evidence. That is, the agent may have evidence about some atomic $\{w\} \in W$, but also about some more complex set $A \subseteq W$. While these pieces of evidence are assumed compatible, they may be incomplete — and DS theory mirrors this incompleteness. In fact, DS theory is specifically targeted at cases where agents receive some imprecise evidence that points to some A without implying any of the atoms in A. For it is only when such evidence exists, i.e. when $m(A) > 0$ for some non-singleton $A \subseteq W$, that DS-belief violates the axioms of classic probability theory.

In non-standard probability functions, on the other hand, the main evidence available to agents concerns atomic propositions p or $\neg p$ and conjunctions thereof. It is this information alone that determines the measure μ in a (canonic) probabilistic model. Information about any logically more

complex propositions can then be derived from this simple information, as it can be read off from the corresponding canonical model. The agents' available evidence on literals, however, may not only be incomplete but also inconsistent, as witnessed by allowing for gaps and gluts.

Besides their informal differences in approach, we can also formally show DS theory and non-standard probability functions to be orthogonal perspectives. For the following theorem, recall that DS theory must assign belief one to classical tautologies and zero to classical contradictions.

Theorem 2 *Assume that* At *is finite and that a non-standard belief function* P *satisfied* $P(\varphi) = 1$ *for all classical tautologies and* $P(\varphi) = 0$ *to all classical contradictions. Then* P *is a classical probability function.*

Proof. Kolmogorov's non-negativity axiom already holds for non-standard probability functions. His normalization axiom, that $P(\varphi) = 1$ for any tautology holds by assumption. As $\mathcal{L}$ is finite, σ-additivity is the same as finite additivity. For this, it suffices that $P(\varphi) + P(\psi) = P(\varphi \vee \psi)$ whenever φ and ψ are mutually exclusive in classical logic. This follows from the import export rule, $P(\varphi) + P(\psi) = P(\varphi \vee \psi) + P(\varphi \wedge \psi)$ together with the fact that $P(\varphi \wedge \psi) = 0$, as $\varphi \wedge \psi$ is a classical contradiction.

Theorem 2 entails that (non-probabilistic) beliefs expressed in terms of DS theory and non-standard probabilities can not be translated into each other. The question remains whether we can elicit some weaker relationship between the two frameworks such as a partial translation function that at least connects a subset of the agent's belief to a subset of her non-standard probabilities. In doing so, we are especially interested whether the learning mechanisms induced by various conditioning policies for DS-functions (FAGIN and HALPERN 1990) and non-standard probability functions (KLEIN, MAJER, and RAFIEE RAD 2020, Section 7) cohere. We start by defining a *partial* correspondence between the two frameworks.

2.2 JOINT SEMANTICS AND A PARTIAL BRIDGE

Despite their incompatibility (cf. Theorem 2) we want to establish at least a partial bridges between DS Theory and non-standard probabilities. To do so, we identify special cases where DS belief is the same as non-standard probability for a large class of formulas. The construction to come largely builds on the semantics for non-standard probability and Dempster-Shafer theory given in Sections 1.1 and 1.2.2 and in particular on Definitions 1 for DS theory and Definition 5 for non-standard probability. In the remainder of

this section, we will define a correspondence map between a certain class of canonical probabilistic models and models for DS theory. In the next section we will then use this map to compare various information updates on either framework that are all generalizations of Bayes' udpate.

In constructing a partial semantic bridge between the two frameworks, we capitalize on the observations made in the previous section. Firstly, we have seen that DS Theory requires beliefs to be non-contradictory. Moreover the requirement of $\mathrm{bel}(W) = 1$ rules out uncertainty about classical tautologies. Likewise, classical contradictions must receive a weight of zero. We can translate these properties into additional requirements on non-standard belief functions and their corresponding canonical probabilistic models (by means of Theorem 1).

Lemma 1 *Let* $P : \mathcal{L} \to [0;1]$ *be a non-standard belief function and* $\mathcal{M} = \langle \mathcal{P}(\mathrm{Lit}), v^+, v^-, \mu \rangle$ *be the canonical probabilistic model such that* $P_{\mathcal{M}} = P$ *. Then*

i) P *assigns zero weight to any classical contradiction iff* $\mu(x) = 0$ *for all* $x \in \mathcal{P}(\mathrm{Lit})$ *that contain both* q *and* $\neg q$ *for some* $q \in \mathrm{At}$ *.*

ii) $P(\bigvee_{a \in \mathrm{Lit}} a) = 1$ *iff* [6] $\mu(\varnothing) = 0$

Proof. *i)* We start with the left to right part. Recall that in the canonical probabilistic model $\mathcal{M} = \langle \mathcal{P}(\mathrm{Lit}), v^+, v^-, \mu \rangle$, the measure μ is a classical probability function over $\mathcal{P}(\mathrm{Lit})$. Towards a contradiction, assume that μ assigns positive weight to some $x \in \mathcal{P}(\mathrm{Lit})$ that contain both q and $\neg q$ for some $q \in \mathrm{At}$. Then for such x, we have

$$P(\bigwedge_{l \in x} l) = P_{\mathcal{M}}(\bigwedge_{l \in x} l) = \sum_{x \subseteq y} \mu(y) \geq \mu(x) > 0.$$

However, the formula $\bigwedge_{l \in x} l$ is a conjunction containing $q \wedge \neg q$ and therefore a classic contradiction. Thus, p assigns positive weight to a classical contradiction which gives the required contradiction. For the converse direction we use the following well-known fact

Fact 1 *If* φ *is a classical contradiction and* ψ *is not then* $\psi \nvDash_L \varphi$ *.*

Let φ be a classical contradiction. Using the above fact and the definition of the valuation, we have

---

[6] Note that $\varnothing$ here stands for the empty subset of Lit, as a state of $\mathcal{M}$.

$$P_{\mathcal{M}}(\varphi) = \sum_{\{x \in \mathcal{P}(\mathrm{Lit}) \mid \bigwedge_{p \in x} p \vDash_L \varphi\}} \mu(x) \le \sum_{\{x \in \mathcal{P}(\mathrm{Lit}) \mid \exists p \in At: p, \neg p \in x\}} \mu(x) = 0,$$

which gives the left hand side as required.

For ii) First remember that Belnap-Dunn logic has no tautologies, but the top-element of its Lindenbaum algebra is $\bigvee_{a \in Lit} a$. Thus for $x \subseteq \mathrm{Lit}$ non-empty we have that $\bigwedge_{l \in x} l \vDash_L \bigvee_{a \in \mathrm{Lit}} a$. Hence, by the definition of canonical probabilistic models,

$$P_{\mathcal{M}}(\bigvee_{a \in \mathrm{Lit}} a) = \sum_{x \subseteq \mathrm{Lit},\, x \neq \varnothing} \mu(x).$$

The right hand side is one exactly if $\mu(\varnothing) = 0$.

Building on the insights from Lemma 1, we define.

Definition 6 A canonical probabilistic model $\mathcal{M} = \langle \mathcal{P}(\mathrm{Lit}), v^+, v^-, \mu \rangle$ is *glut free* iff $\mu(x) = 0$ for all $x \in \mathcal{P}(\mathrm{Lit})$ that contain both q and $\neg q$ for some $q \in At$. Moreover, $\mathcal{M}$ is *normal* iff it is glut free and satisfies $\mu(\varnothing) = 0$.

On the class of normal models we can establish a relation to Dempster-Shafer theory. To do so, we will define a partial map ι from the class of normal probabilistic models to DS models and show that for many formulas φ, the non-standard probability $P_{\mathcal{M}}(\varphi)$ is the same as $\mathrm{bel}(\varphi)$ calculated in $\iota(\mathcal{M})$.

To define this map ι, let Val be the set of valuations over At, i.e.

$$Val = \{x \subseteq \mathrm{Lit} \mid \forall p \in At : p \in x \Leftrightarrow \neg p \notin x\}.$$

We will define DS models over $W = \mathrm{Val}$. That is, each DS model will be of the form (Val, m), where $m : \mathcal{P}(\mathrm{Val}) \to [0,1]$ is a mass function satisfying $\sum_{x \in \mathcal{P}(\mathrm{Val})} m(x) = 1$ and $m(\varnothing) = 0$. Before we proceed, note that the induced belief function bel then assigns a belief to each subset A of Val. Representing $A \subseteq \mathrm{Val}$ as $\bigvee_{x \in A} \bigwedge_{l \in x} l$, we can identify each $A \subseteq Val$ with a corresponding formula. A fortiori, every formula in $\mathcal{L}$ is (classically) logically equivalent to a formula of the form $\bigvee_{x \in B} \bigwedge_{l \in x} l$ where $B \subseteq \mathrm{Lit}$. Hence, every DS-belief function over Val can be identified with a *DS*-belief function over $\mathcal{L}$.

We now construct the map ι that is defined on the class of normal canonical models.

Definition 7 For a normal canonical model $\mathcal{M} = \langle \mathcal{P}(Lit), v^+, v^- \mu \rangle$, the DS-model $\iota(\mathcal{M}) = (Val, \iota^*(\mathcal{M}))$ is defined as

$$\iota^*(\mathcal{M})(A) = \begin{cases} \mu(y) & \text{iff there is a (unique) } y \subseteq \text{Lit with } \vDash \bigwedge_{l \in y} l \leftrightarrow \bigvee_{x \in A} \bigwedge_{l \in x} l \\ 0 \text{ else} \end{cases}$$

In other words, $m(\{x\})$ for $x \in \text{Val}$ is the same as $\mu(x)$ (exploiting normality and that $Val \subseteq \text{Lit}$), while $m(\{x \subseteq \text{Lit} \mid p, q \in x\}$ is the same as $\mu(p \wedge q)\ldots$. Note that the definition implies that $\iota^*(\mathcal{M})(A) = 0$ whenever A is *not* of the form $\{x \in Val \mid y \subseteq x\}$ for some $y \subseteq \text{Lit}$.

As μ was a probability function, $\mu(\varnothing) = 0$ and $\mu(x) = 0$ whenenver $\bigwedge_{l \in x} l$ is a classic contradiction, we obtain the following fact

Fact 2 *For each $x \subseteq Lit$ with $\mu(x) > 0$, there is a unique $A \subseteq \text{Val}$, defined by $A := \{y \in \text{Val} \mid x \subseteq y\}$ with $\mu(x) = m(A)$.*

As μ was a probability function, this implies that $\iota^*(\mathcal{M})$ is a DS mass function on Val. That is, $\iota(\mathcal{M}) = (\text{Val}, \iota^*(\mathcal{M}))$ is indeed a DS model. We denote the corresponding belief function by $\text{bel}_{\iota(\mathcal{M})}$. After these preliminaries, we are now in a position to state our central connection result.

Theorem 3 *Let $\mathcal{M} = \langle \sum, v^+, v^-, \mu \rangle$ be a normal canonical model and $\iota(\mathcal{M}) = (\text{Val}, \iota^*(\mathcal{M}))$ be its corresponding Dempster-Shafer model. Then for any $\varphi \in \mathcal{L}$ that is a conjunction of literals, $P_{\mathcal{M}}(\varphi) = \text{bel}_{\iota(\mathcal{M})}(\varphi)$*

In other words, the function ι establishes a partial connection between non-standard probability functions and belief functions: the probability of a conjunction of literals φ is the same as the belief in φ in the DS model obtained from ι as above.

Proof. First, note that for any $y \in \text{Lit}$ and any φ that is a conjunction of a set of literals $L(\varphi)$, we have

$$\bigwedge_{l \in y} l \vDash_L \varphi \Leftrightarrow L(\varphi) \subseteq y$$

Thus, for probabilities, we have that

$$P_{\mathcal{M}}(\varphi) = \sum_{y \supseteq L(\varphi)} \mu(y).$$

Likewise, we have for any valuation $v \in Val$ that $v \vDash \varphi$ iff $L(\varphi) \subseteq v$. Thus, in $\iota(\mathcal{M})$, we have

$$\mathrm{bel}_{\iota(\mathcal{M})}(\varphi) = \sum_{\{X \subseteq \mathrm{Val} | \forall x \in X : x \supseteq A\}} \iota^*(\mathcal{M})(X).$$

By the construction of $\iota(\mathcal{M})$, we have

$$\sum_{\{X \subseteq \mathrm{Val} | \forall x \in X x \supseteq A\}} \iota^*(\mathcal{M})(X) = \sum_{y \supseteq A} \mu(y),$$

finishing the proof.

It is important to note that we cannot hope for a stronger correlation between the two frameworks. By Theorem 2, the functions $P_{\mathcal{M}}$ and $\mathrm{bel}_{\iota(\mathcal{M})}$ can only coincide on *all* formulas when $P_{\mathcal{M}}$ is a *classic* belief function. This, in turn, is the case iff $\mu(x) = 0$ whenever $x \not\subseteq \mathrm{Val}$.

Nevertheless, we will employ the partial correspondence function ι for a cautious comparison of the two framework in terms of the update dynamics induced by different conditionaliziation strategies.

3. DYNAMICS OF LEARNING:
NON-STANDARD PROBABILITIES VS DS MODELS

When interpreting the non-standard probabilities introduced in the previous sections as some agent's (non-standard) degrees of information-based belief, a natural next step is to inquire into belief dynamics. That is, we may ask how agents should ideally update their information upon receiving new information.

Incorporating new information has emerged as a central topic within a wide range of disciplines. It is, for instance, at the core of Bayesian epistemology and by extension Bayesian statistics and Bayesian inference. It also constitutes significant portions of the current literature on Belief Revision (ALCHOURRÓN, GÄRDENFORS, and MAKINSO. 1985; KATSUNO and MENDELZON 1991), Decision Theory (KLEIN, MARX, and SCHELLER 2018), Social Choice Theory or expert systems. Moreover, information updates also play a crucial role in AI and design and analysis of efficient machine learning algorithms (BUSH and MOSTELLER 1955; EREV and ROTH 1998). Although there is a vast amount of literature on these topics, the core of Bayesian learning is a simple rule from classical probability theory — Bayesian conditioning.

Classical Bayesian learning concerns a scenario where an agent learns that some proposition φ she was uncertain about (her subjective belief was strictly between 0 and 1) holds in fact true. After receiving this information

the agent obviously has to assign probability one to φ. On top, learning that φ may have implications for other probability judgments, so she has to recalculate all remaining probabilities accordingly.

Semantically we can see this situation as one of changing the distribution of probability measure between $|\varphi|$ and $|\neg\varphi|$. That is to change the distribution of probability between the set of φ-worlds and non-φ worlds (which determine a binary partition of the set of possible worlds), cf. Figure 1. After learning that φ is true the agent changes the old probability $\mu(|\varphi|)$ of the partition $|\varphi|$ to an updated probability $\mu'(|\varphi|) = 1$ and the probability of the partition $|\neg\varphi|$ to 0. Naturally, it follows that the probability of any subset of $|\neg\varphi|$, i.e. any intersection between $|\neg\varphi|$ and some $|\psi|$ is also zero. Hence, for each proposition ψ, only the part of its extension $|\psi|$ that intersects $|\varphi|$, i.e. $|\psi \wedge \varphi|$ maintains a positive probability after the update. These updated values are then normalized to obtain the updated probability function. Formally, this is specified by the formula
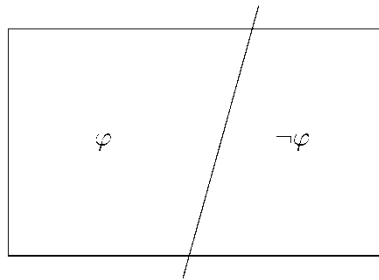

Figure 1. Classic Conditioning

$$\mu'(|\psi|) = \frac{\mu(|\psi \wedge \varphi|)}{\mu(|\varphi|)},$$

where μ' is the updated measure after learning that φ.

Bayes conditioning, however, is limited in the sense that it allows only for learning of certain types of information. In some situations an agent might not learn ϕ for certain, but only a new (presumably more accurate) estimation for its probability, say $\mu'(\phi) = q$. This is addressed by Jeffery conditionalization, a generalization of Bayesian conditioning. In this case, the agent dos not simply throw away the partition $|\neg\phi|$ be reducing its probability to zero. Instead she has to expand or contract the original measure μ proportionally on $|\phi|$ and $|\neg\phi|$ in a way that $\mu'(|\phi|) = q$ and $\mu'(|\neg\phi|) = 1 - q$. We hence get for any formula $\psi \in \mathcal{L}$:

$$\mu'(\psi) = \mu(|\ \psi \wedge \phi\ |)\frac{q}{\mu(|\ \phi\ |)} + \mu(|\ \psi \wedge \neg\phi\ |)\frac{1-q}{\mu(|\ \neg\phi\ |)}.$$

In the case of learning the new information with certainty (i.e. $q = 1$), Jeffrey updating reduces to the formula for Bayes' conditioning.

Just as within the classical setting, the question on how to incorporate the new information that φ can also be asked with respect to the weakened frameworks discussed here, non-standard probability theory and Dempster-Shafer belief functions. In the present section, we compare these two frameworks with respect to their learning and updating rules. In doing so, we restrict ourselves to counterparts of Bayes updating, i.e. to situations where φ is learned for certain.

Klein, Majer, and Rafiee Rad (2020) provide a detailed study of conditionalization for non-standard probabilities. There, conditionalization of non-standard probabilities is approached by generalizing Jeffery conditioning to the non-standard setting, from which Bayesian conditionalization can be defined as a special case. In fact, the generalization of Jeffrey conditioning does not turn out unique. There are two natural generalizations of Jeffrey updating that give rise to five different notions of Bayes update. For the discussion to follow, we briefly introduce what are arguably the most straightforward generalizations of Jeffrey update and Bayes conditioning.

Consider an agent in a non-standard probabilistic framework who learns a new probability, say $q \in [0,1]$, for some φ (whose prior probability was in $(0,1)$). Importantly, setting the posterior probability of ϕ to be q does not carry any implications regarding the probability of $\neg\phi$ within our non-standard framework. For instance, the agent may or may not decide to leave $P(\neg\varphi)$ unchanged in her update.

Let us first consider non-standard Jeffrey updating semantically. For any $\phi \in \mathcal{L}$, we can dissect the state space of a probabilistic model $\mathcal{M} = \langle \Sigma, \mu, v^+, v^- \rangle$ into two sets – the truth set $|\ \varphi\ |$ of φ and it's complement $\Sigma \setminus |\ \varphi\ |$. Unlike in the classic case, however, $\Sigma \setminus |\ \varphi\ |$ is not the truth set of $|\ \neg\varphi\ |$ – nor of any other formula. Following the intuition for classical case, the non-standard Jeffrey updating is defined as follows.

Definition 8 Let $\mathcal{M} = \langle \Sigma, v^+, v^-, \mu \rangle$ be a probabilistic model and let $q \in [0,1]$ and $\varphi \in \mathcal{L}$ such that $\mu(|\ \varphi\ |) \in (0,1)$. Then the semantic non-standard Jeffrey update of $\mathcal{M}$ for updating the probability of φ to q is the probabilistic model $\mathcal{M}^{\varphi,q} = \langle \Sigma, v^+, v^-, \mu^{\varphi,q} \rangle$ determined by:

$$\mu^{\varphi,q}(\{x\}) = \begin{cases} \mu(\{x\}) \cdot \dfrac{q}{\mu([\varphi])} & \text{iff } x \in [\varphi] \\[2ex] \mu(\{x\}) \cdot \dfrac{1-q}{1-\mu([\varphi])} & \text{else.} \end{cases}$$

It is easy to check that non-standard Jeffrey updating is successful. That is for any probabilistic model $\mathcal{M} = \langle \Sigma, v^+, v^-, \mu \rangle$, $\varphi \in \mathcal{L}$ and $q \in [0,1]$ with $\mu(|\varphi|) \neq 0$ the updated model $\mathcal{M}^{\varphi,q}$ satisfies $\mu^{\varphi,q}(x)([\varphi]) = q$.

As detailed in Klein, Majer, and Rafiee Rad (2020), non-standard Jeffrey-updating can be defined purely syntactically.

Definition 9 Let $p : \mathcal{L} \to \mathbb{R}$ be a non-standard probability assignment, let $q \in [0;1]$ and $\varphi \in \mathcal{L}$ with $P(\varphi) \in (0;1)$. Then the syntactic non-standard Jeffrey update setting the probability of φ to q is the probability function $P^{\varphi,q} : \mathcal{L} \to \mathbb{R}$ defined by

$$P^{\varphi,q}(\psi) = P(\psi \wedge \varphi) \cdot \frac{q}{P(\varphi)} + (P(\psi) - P(\psi \wedge \varphi))\frac{1-q}{1-P(\varphi)}$$

Just as in the classic case, Klein, Majer, and Rafiee Rad (2020) define non-standard Bayesian updating as special case of non-standard Jeffrey updating where the probability of φ is set to 1. In this case, the formula of Definition 9 simplifies to the same formula as in the classical case.

Definition 10 Let $p : \mathcal{L} \to \mathbb{R}$ be a non-standard probability function and let $\varphi \in \mathcal{L}$ with $P(\varphi) > 0$. Then the positive non-standard Bayesian update on φ (corresponding to the information that φ is true) is the function $P(\psi \mid \varphi)$:

$$P(\psi \mid \varphi) = \frac{P(\psi \wedge \varphi)}{P(\varphi)} \quad \text{for} \varphi \in \mathcal{L}.$$

With this construction, Jeffery conditioning for non-standard probabilities and its special case of non-standard Bayesian conditioning, follow the exact same rationale as in the classic case. In this sense the non-standard Jeffery and non-standard Bayesian conditionings remain fully faithful to the classical intuition. It follows immediately, that $P(\psi \mid \varphi)$ amounts to classic Bayes conditioning whenever the prior non-standard belief function P satisfies Kolmogorov's axioms.

Moreover, Bayes Theorem, central to Bayesian inference and Bayesian learning, continues to hold for positive non-standard Bayesian updates:

Lemma 2 *Let* φ, ψ *be such that* $P(\varphi) > 0$. *Then*

$$P(\psi \mid \varphi) = P(\varphi \mid \psi) \cdot \frac{P(\psi)}{P(\varphi)}$$

Notably, however, non-standard Bayes conditioning is less expressive than classic conditioning in the following sense. Within a classic setting, learning some φ to be *false*, i.e. setting its probability to 0, is the same as learning the probability of $\neg\varphi$ to be 1. Hence, no separate mechanism for learning about the falsity of some φ is needed. This ceases to hold in the non-standard setting. There, the propositions that $P(\varphi) = 0$ and that $P(\neg\varphi) = 1$ are mutually independent. Hence learning that φ is false cannot be reduced to learning about the truth of $\neg\varphi$ and a separate rule is needed for learning about the falsity of propositions.[7]

Also for Dempster-Shafter belief functions, several updating rules have been proposed.[8] The main two proposals, which we shall consider here, are due to Dempster (best illustrated in SHAFER 1976), and to Fagin and Halpern (1990). Crucially, both are generalizations of Bayesian conditioning in that they correspond to Bayes updating when the underlying belief function is a classical probability measure. The two rules are, however, known to diverge in general Fagin and Halpern (1990). To be more precise, Dempster (1976) defines conditioning of a belief function bel by

$$\mathrm{bel}_D(\varphi \mid \psi) := \frac{\mathrm{bel}(\varphi \vee \neg\psi) - \mathrm{bel}(\neg\psi)}{1 - \mathrm{bel}(\neg\psi)}.$$

The second strategy, proposed in Fagin and Halpern (1990) is originally motivated semantically. Its underlying idea is to represent a Dempster-Shafer belief function as the lower limit of a family of belief functions - and update each of these belief functions separately. This rule, however, can be identified in a purely syntactic manner as shown below. For more details see Fagin and Halpern (1990).

$$\mathrm{bel}_{FH}(\varphi \mid \psi) := \frac{\mathrm{bel}(\varphi \wedge \psi)}{\mathrm{bel}(\varphi \wedge \psi) + (1 - \mathrm{bel}(\varphi \vee \neg\psi))}.$$

---

[7] This rule is discussed in Klein, Majer, and Rafiee Rad (2020), alongside with further rules for updating on conflict or uncertainty.

[8] We should emphasize that all rules discussed here are proposed as *definitions* of updating policies. Building on these one may construct measures of confirmation that generalize the various Bayesian confirmation measures (HARTMANN and SPRENGER 2019) to nonstandard settings. We leave this for future work.

In what follows, we use the map ι of the previous section to compare the three updating rules. While $\mathrm{bel}_D(\varphi \mid \psi)$ and $\mathrm{bel}_{FH}(\varphi \mid \psi)$ are known to disagree in general, we can ask whether they agree on $\iota(\mathcal{M})$ for a normal canonical model $\mathcal{M}$. There, we can also ask whether either of these agree with the corresponding conditional function $P_{\mathcal{M}}(\varphi \mid \psi)$ derived from $\mathcal{M}$, at least when φ and ψ are both conjunctions of literals (and hence Theorem 3 applies).

In fact, given that $P_{\mathcal{M}}(\varphi \mid \psi)$ is defined by a straightforward generalization of Bayes conditioning to non-standard probabilities, one may arguably use the question of whether it yields the same values as $\mathrm{bel}_D(\varphi \mid \psi)$ or $\mathrm{bel}_{FH}(\varphi \mid \psi)$ in $\iota(\mathcal{M})$ to assess the adequacy of the latter two, at least when φ and ψ are as in Theorem 3. The answer, however, is negative. The three conditional probabilities (resp. beliefs) $P_{\mathcal{M}}(\varphi \mid \psi)$ (in $\mathcal{M}$) and $\mathrm{bel}_D(\varphi \mid \psi)$, $\mathrm{bel}_{FH}(\varphi \mid \psi)$ (in $\iota(\mathcal{M})$ may already come apart in simple cases where $\mathcal{M}$ is a normal canonical model and φ, ψ are atomic.

Example. Let $\mathrm{At} = \{p, q\}$ and consider the normal canonical model $\mathcal{M} = \langle \mathcal{P}(\mathrm{Lit}), v^+, v^- \mu \rangle$ where μ is defined by

$$\mu(x) = \begin{cases} \dfrac{1}{8} & \text{iff } x \in \{p, \neg p, q, \neg q, p \wedge q, p \wedge \neg q, \neg p \wedge q, \neg p \wedge \neg q\} \\ 0 & \text{else.} \end{cases}$$

Then $\iota(\mathcal{M}) = (\mathrm{Val}, \iota^*(\mathcal{M}))$ is defined through the weight function

$$\iota^*(\mathcal{M}) : \mathcal{P}(\{p \wedge q, p \wedge \neg q, \neg p \wedge q, \neg p \wedge \neg q\}) \to [0;1]$$

given by

$$\iota^*(\mathcal{M})(A) = \begin{cases} \dfrac{1}{8} & \text{iff } A \in \{\{p \wedge q\}, \{p \wedge \neg q\}, \{\neg p \wedge q\}, \{\neg p \wedge \neg q\}, \{p \wedge q, p \wedge \neg q\}, \\ & \quad \{\neg p \wedge q, \neg p \wedge \neg q\}, \{p \wedge q, \neg p \wedge q\}, \{p \wedge \neg q, \neg p \wedge \neg q\}\} \\ 0 & \text{else.} \end{cases}$$

With this, we calculate $P(p \mid q)$ on $\mathcal{M}$ and $\mathrm{bel}_D(p \mid q)$, $\mathrm{bel}_{FG}(p \mid q)$ on $\iota(\mathcal{M})$. We start with $P(p \mid q)$. Taking into account that the truth sets of $p \wedge q$ and q in $\mathcal{M}$ are $\{\{p, q\}\}$ and $\{\{p, q\}, \{\neg p, q\}, \{q\}\}$ respectively and that each of these states receives the same weight of $\dfrac{1}{8}$, we obtain that

$$P(p \mid q) = \frac{P(p \wedge q)}{P(q)} = \frac{1}{3}$$

For the two conditional belief functions we move to $\iota(\mathcal{M})$. First, we calculate $\mathrm{bel}_D(p \mid q)$. By definition,

$$\mathrm{bel}(p \vee \neg q) = \sum_{\{A \subseteq \{p \wedge q, p \wedge \neg q, \neg p \wedge \neg q\}\}} w(A) = \frac{5}{8}$$

$$\mathrm{bel}(\neg q) = \sum_{\{A \subseteq \{p \wedge \neg q, \neg p \wedge \neg q\}\}} w(A) = \frac{3}{8}$$

Hence

$$\mathrm{bel}_D(p \mid q) := \frac{\mathrm{bel}(p \vee \neg q) - \mathrm{bel}(\neg q)}{1 - \mathrm{bel}(\neg q)} = \frac{\frac{5}{8} - \frac{3}{8}}{1 - \frac{3}{8}} = \frac{2}{5}$$

Lastly, for $\mathrm{bel}_{FH}(p \mid q)$ we recall that $\mathrm{bel}(p \wedge q) = \sum_{\{A \subseteq \{p \wedge q\}\}} w(A) = \frac{1}{8}$ and thus

$$\mathrm{bel}_{FH}(p \mid q) := \frac{\mathrm{bel}(p \wedge q)}{\mathrm{bel}(p \wedge q) + (1 - \mathrm{bel}(p \vee \neg q))} = \frac{\frac{1}{8}}{\frac{1}{8} + 1 - \frac{5}{8}} = \frac{1}{4}.$$

Hence, the three updating rules differ on the conditional probability (resp. belief) in p after learning q.

As pointed out above the conditionalization of non-standard probabilities $P(p \mid q)$ directly imports the intuition behind classical conditionalization to the non-standard setting. With this perspective the disagreement between $P(p \mid q)$ and the conditionalization proposed by Dempster or Fagin and Halpern can arguably be taken to indicate that either i) the map $\iota(\mathcal{M})$ is too weak to preserve conditioning, even though $P(p) = \mathrm{bel}(p)$, $P(q) = \mathrm{bel}(q)$ and $P(p \wedge q) = \mathrm{bel}(p \wedge q)$ or ii), that bel_D and bel_{FH} operate on intuitions that are different from those of classical conditionalization, and in this sense, arguably deviate from the spirit of Bayes update.

4. CONCLUSION

We studied and compared two frameworks for quantitative uncertain reasoning in situations where the available information may be inconsistent

and incomplete. The first of these is based on a non-standard probability theory that assigns a system of weights satsifying weaker axioms than those of Kolmogorov. This non-standard probability, in term, is built on the para-consistent Belnap-Dunn logic, a well studied logic for reasoning about imperfect information. Moving to BD as underlying system generalises the classical approach in two direction. First, it allows for the underlying information to be incomplete or even inconsistent, thus relaxing the assumption of information completeness. Second, it permits to draw conclusions from imperfect bodies of information by using an entailment system that does not validate explosion and where classic tautologies such as $\phi \vee \neg \phi$ cease to be logical validities. Both these generalisations are crucial for reasoning with and about uncertain and possibly faulty information. In situations of informational scarcity, for instance agents will likely not receive sufficient information for or against some proposition to even forme a reliable classical likelihood judgment. Likewise, when evidence is gathered from various sources, it may plausibly contain conflicting or contradictory information. As Belnap-Dunn logic has been developed to address both aspects, it seems suitable as underlying logic for uncertain reasoning in quantitative scenarios. In fact, probabilistic extensions of Belnap Dunn logic have recently been studied in detail by Childers, Majer, and Milne (2019) and Klein, Dominik, Majer, Rafiee Rad (2020).

Within the literature on uncertain reasoning, however, generalizations of BD logic are neither the only nor the first approaches towards representing uncertainty in quantitative contexts. One of the most prominent of approaches in this literature arguably is Dempster-Shafer theory, defining evidence based Dempster-Shafer belief functions. The main objective of this paper was to compare non-standard probability theory and Dempster-Shafer belief functions in their representation of uncertain reasoning

A first distinction between the two frameworks suggested itself rather immediately. While the probabilistic extension of BD logic allows for handling both inconsistent and incomplete information, Dempster-Shafer belief functions can only deal with the latter imperfection. In particular, Dempster-Shafer theory can neither assign non-zero belief to a classic contradiction, nor have belief values for two contradictory statements that add up to more than 1. On the face of it, this already suggest that the non-standard probability setting may be more general. As we have shown, however, this is not exactly correct. While the non-standard probability framework fairs better in dealing with inconsistencies, Dempster-Shafer belief functions allow for im-

precise evidence or local gluts where $\mathrm{bel}(A) + \mathrm{bel}(B) < \mathrm{bel}(A \vee B)$, which are ruled out for non-standard probabilities.

Thus, none of these frameworks is more general than the other. And indeed, we have shown the two frameworks to be orthogonal in the types of imperfections they can represent. More specifically, the only quantitative belief (resp. probability) ascriptions that are judged admissible by both Dempster-Shafer theory and DB-based non-standard probability are classic probability functions — which are devoid of any of the imperfections discussed.

As next step we then investigated how existing learning or updating mechanisms in both framework compare. That is, upon starting from the same point in different frameworks, we asked whether or not the respective conditioning rules agree. To be more precise let (Σ', m) be a Demspter Shafer model with the induced belief function bel, and (Σ, v^+, v^-, μ) a non-standard probability model with the induced non-standard probability p, and $\phi, \psi \in L$ such that $P(\phi) = \mathrm{bel}(\phi)$, $P(\psi) = \mathrm{bel}(\psi)$ and $P(\phi \wedge \psi) = \mathrm{bel}(\phi \wedge \psi)$. Our question, then, was whether conditionalization of non-standard probability agrees with either of the two conditionalization policies proposed for Dempster-Shafer belief functions. As we have shown this already fails in very simple cases.

We take this as a hint, that the updating rules must differ substantially in their underlying assumptions, even though they all boil down to Bayes' updating when applied to classic probability assignments. Moreover, given that conditionalization for non-standard probabilities is defined in exactly the same way as for classic probability theory, it can arguably be seen as a natural generalization of the classical Bayesian intuitions to non-classical settings. Accordingly, if we take agreement with non-standard conditionalization as a measure of faithfulness to the classical Bayesian intuitions, the fact that both DS conditioning rules behave differently might suggest that they are, in some sense, significantly different to classical conditionalization.

BIBLIOGRAPHY

ALCHOURRÓN, Carlos E., Peter GÄRDENFORS, and David MAKINSON. 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic* 50, No. 2: 510–530.

ANDERSON, Alan Ross, and Nuel D. BELNAP. 1975. *Entailment: The Logic of Relevance and Necessity*. Princeton, NJ: Princeton University Press.

BALTAG, Alexandru, Bryan RENNE, and Sonja SMETS. 2014. "The Logic of Justified Belief, Explicit Knowledge, and Conclusive Evidence. *Annals of Pure and Applied Logic* 165, issue 1: 49–81.

BALTAG, Alexandru, Lawrence S. MOSS, and Sławomir SOLECKI (1998). "The Logic of Common Knowledge, Public Announcements and Private Suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, edited by Itzhak Gilboa, 43–56. San Francisco, CA: Morgan Kaufmann Publishers.

BATENS, Diderik. 2001). "A General Characterization of Adaptive Logics." *Logique et Analyse* 44, No. 173-175: 45–68.

BELNAP, Nuel D. 1977. "A Useful Four-Valued Logic." *Modern Uses of Multiple-Valued Logic: invited papers from the Fifth International Symposium on Multiple-Valued Logic, held at Indiana University, Bloomington, Indiana, May 13-16, 1975*, edited by J. Michael Dunn and George Epstein, 5–37. Dordrecht, Boston: D. Reidel.

BELNAP, Nuel D. 2019. "How a Computer Should Think." In *New Essays on Belnap-Dunn Logic*, edited by Hitoshi Omori and Heinrich Wansing, 35–53. Springer International Publishing.

VAN BENTHEM, Joahan, Jan VAN EIJCK, and Barteld KOOI. 2006. "Logics of Communication and Changet." *Information and Vomputation* 204, No. 11: 1620–1662.

VAN BENTHEM, Johan. 2011. *Logical Dynamics of Information and Interaction*. Cambridge, MA: Cambridge University Press

BUSH, Robert R. and Frederick MOSTELLER. 1955. *Stochastic Models for Learning*. Hoboken, NJ: Wiley.

CHILDERS, Timothy, Ondrej MAJER, and Peter MILNE. 2019. *The (Relevant) Logic of Scientific Discovery* (under review).

DA COSTA, Newton C.A. 1974. "On the Theory of Inconsistent Formal Systems." *Notre Dame Journal of Formal Logic* 15, No. 4: 497–510.

DA COSTA, Newton C.A., and V. SUBRAHMANIAN. 1989. "Paraconsistent Logic as a Formalism for Reasoning about Inconsistent Knowledge Bases." *Artificial Intelligence in Medicine* 1, issue 4: 167–174. DOI: https://doi.org/10.1016/0933-3657(89)90031-6.

DARWICHE, Adnan, and Judea PEARL. 1997. "On the Logic of Iterated Belief Revision." *Artificial Intelligence* 89, issue 1–2: 1–29. doi: 10.1016/S0004-3702(96)00038-0.

DEMPSTER, Arthur P. 1967. "Upper and Lower Probabilities Induced by a Multivalued Mapping." *The Annals of Mathematical Statistics* 38, No. 2, 325–339.

VAN DITMARSCH, Hans, Wiebe VAN DER HOEK, Barteld KOOI. 2008. "Dynamic Epistemic Logic." *Synthese* 337.

DUNN, J. Michael, and Nicholas M. KIEFER. 2019. "Contradictory Information: Better than Nothing? The Paradox of the Two Firefighters." In *Graham Priest on Dialetheism and Paraconsistency*, edited by Can Başkent, and Thomas Macaulay Ferguson, 231–247. Springer International Publishing.

DUNN, J. Michael. 1976. "Intuitive Semantics for First Degree Entailment and 'Coupled Trees'." *Philosophicl Studies* 29, No. 3: 149–168.

DUNN, J. Michael. 2010). Contradictory Information: Too Much of a Good Thing. *Journal of Philosophical Logic 39*(4), 425–452.

EREV, Ido, and Alvin E. ROTH. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review* 88, No. 4: 848–881.

FAGIN, Ronald, and Joseph Y. HALPERN. 1990. "A New Approach to Updating Beliefs." In *UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence* edited by Piero P. Bonissone, Max Henrion, Laveen N. Kanal, John F. Lemmer, 347–374. Elsevier.

FAGIN, Ronald, and Joseph Y. HALPERN. 1991. "Uncertainty, Belief, and Probability." *Computational Intelligence* 7, No. 3: 160–173.

FAGIN, Ronald, Joseph Y. HALPERN, and Nimrod MEGIDDO. 1990. "A Logic for Reasoning about Probabilities." *Information and Computation* 87, issue 1–2: 78–128. doi: 10.1016/0890-5401(90)90060-U.

HALPERN, Joseph Y. 2017. *Reasoning about Uncertainty*. Cambridge, MA: MIT Press.

JAŚKOWSKI, Stanisław. 1969. "Propositional Calculus for Contradictory Deductive Systems." *Studia Logica* 24: 143–157

JØSANG, Audun. 1997. "Artificial Reasoning with Subjective Logic." In *Proceedings of the Second Australian Workshop on commonsense Reasoning*, 48: 34.

KATSUNO, Hirofumi, and Alberto O. MENDELZON. 1991. "Propositional Knowledge Base Revision and Minimal Change." *Artificial Intelligence* 52, No. 3: 263–294.

KLEIN, Dominik, Johanes MARX, and Simon SCHELLER. 2018. "Rationality in context: On Inequality and the Epistemic Problems of Maximizing Expected Utility." *Synthese*. doi: 10.1007/s11229-018-1773-0.

KLEIN, Dominik, Ondrej MAJER, and Sourush RAFIEE RAD. 2020. "Probabilities with Gaps and Gluts" *arXiv:2003.07408.*

KOLMOGOROV, A.N. 2018. *Foundations of the theory of probability*. Courier Dover Publications.

KOOI, Bartel P. 2003. "Probabilistic Dynamic Epistemic Logic." *Journal of Logic, Language and Information* 12: 381–408

LEITGEB, Hannes, and Krister SEGERBERG. 2007, "Dynamic Doxastic Logic: Why, How, and Where To?." *Synthese*, 155: 167–190.

PARIS, J.B. 1994. *The Uncertain Reasoners' Companion: A Mathematical Perspective*. (Cambridge Tracts in Theoretical Computer Science, Series Number 39). Cambridge, MA: Cambridge University Press.

PRIEST, Graham. 1979. "Logic of Paradox." *Journal of Philosophical Logic* 8: 219–241.

PRIEST, Graham. 2006. *In contradiction*. Oxford: Oxford University Press.

PRIEST, Graham. 2007. "Paraconsistency and dialetheism." *Handbook of the History of Logic* 8: 129–204.

RESCHER, Nicholas, and Ruth MANOR. 1970. "Propositional Calculus for Contradictory Deductive Systems." *Theory and Decision* 1, No. 2: 179–217.

SHAFER, Glenn. 1976. *A Mathematical Theory of Evidence*. Vol. 42. Princeton, NJ: Princeton University Press.

SPRENGER, Jan, and Stephan HARTMANN. 2019. *Bayesian Philosophy of Science*. Oxford: Oxford University Press.

ZHOU, Chunlai. 2013). "Belief Functions on Distributive Lattices." *Artificial Intelligence* 201: 1–31.

NON-CLASSICAL PROBABILITIES FOR DECISION MAKING IN SITUATIONS OF UNCERTAINTY

S u m m a r y

Analyzing situations where information is partial, incomplete or contradictory has created a demand for quantitative belief measures that are weaker than classic probability theory. In this paper, we compare two frameworks that have been proposed for this task, Dempster-Shafer

theory and non-standard probability theory based on Belnap-Dunn logic. We show the two frameworks to assume orthogonal perspectives on informational shortcomings, but also provide a partial correspondence result. Lastly, we also compare various dynamical rules of the two frameworks, all seen as generalizations of classic Bayes' conditiong.

Keywords: reasoning under uncertinty; non-classcal probability; Dempster-Shafer theory; Belnap-Dunn logic.

NIEKLASYCZNE PRAWDOPODOBIEŃSTWA NA POTRZEBY PODEJMOWANIA DECYZJI W SYTUACJACH NIEPEWNOŚCI

Summary

Analiza sytuacji, w których informacja jest częściowa, niepełna bądź niespójna wskazuje na potrzebę zbudowania jakościowych miar siły przekonań odmiennych niż te, które są oferowane przez klasyczną teorię prawdopodobieństwa. W niniejszej pracy porównujemy dwa ujęcia zaproponowane dla realizacji tej potrzeby: teorię Dempstera-Shafera i niestandardową teorię prawdopodobieństwa nabudowaną na logice Belnapa-Dunna. Pokazujemy, że te dwa formalizmy przyjmują ortogonalne perspektywy postrzegania niedostatków informacyjnych, a jednocześnie dostarczają rezultatów częściowo ze sobą korespondujących. Na koniec porównujemy różne dynamiczne reguły z obu formalizmów traktując je wszystkie jako uogólnienie warunkowania Bayesowskiego.

Słowa kluczowe: rozumowania w sytuacji niepewności; nieklasyczne prawdopodobieństwo; teoria Dempstera-Shafera; logika Belnapa-Dunna.