Conditionals Based on Selection Functions, Modal Operators and Probabilities

Tommaso Flaminio
IIIA - CSIC
Campus de la UAB
Barcelona, Spain
tommaso@iiia.csic.es

Lluis Godo
IIIA - CSIC
Campus de la UAB
Barcelona, Spain
godo@iiia.csic.es

Giuliano Rosella
Department of Philosophy
University of Turin
Turin, Italy
giuliano.rosella@unito.it

1 Introduction

Methods for probability updating, of which Bayesian conditionalization is the most well-known and widely used, are modeling tools that aim to represent the process of modifying an initial epistemic state, typically represented by a prior probability function P, which is adjusted in light of new information. Notably, updating methods and conditional sentences seem to intuitively share a deep connection, as is evident in the case of conditionalization. Starting with a probability P, the question of what the probability of an event B is, given that A holds (i.e., $P(B \mid A)$), appears to involve conditional reasoning of the form "if A then B". Indeed, Adams [1] argues that the assertability of (indicative) conditionals, represented as "if A then B", aligns with the corresponding conditional probability $P(B \mid A)$. However, this intuitive connection between conditionals and updating methods has been challenged by Lewis's celebrated triviality result [20]. This result demonstrates that conditional probability $P(B \mid A)$ cannot generally be equated with the probability of the corresponding conditional, $P(A \rhd B)$, without leading to trivializing constraints on the initial probability functions.

Building upon these foundational works, research on the relationship between updating methods and conditionals has been and continues to be prolific. This research has primarily focused on understanding whether and how updating methods can be semantically represented in terms of conditional connectives, or conversely, what kind of updating methods are associated with specific semantic conditional operators (see, for instance, [22, 9, 15, 10, 7, 23]).

The present work contributes to this line of research and aims at shedding new light on the relationship between updating methods and conditional connectives. Departing from previous literature that often focused on a specific type of conditional or a particular updating method, our goal is to prove general results concerning the connection between conditionals and their probabilities. This will allow us to characterize the probabilities of certain conditional connectives and to understand what class of updating procedures can be represented using specific conditional connectives. Broadly, we adopt a general perspective that encompasses a large class of conditionals and a wide range of updating methods, enabling us to prove some general results concerning their interrelation.

Let us begin by providing some background. As observed in [14] (see also [9]), conditionalization updating can be situated within the broader context of *imaging* updating methods. Thge imaging method was initially introduced by Lewis [20] as a way to circumvent his triviality result and it hs been extended and generalized by several authors, notably by Gärdenfors [12]. While conditionalization is an updating method that requires only an algebra of events and a (positive) probability function, imaging methods update an a prior probability by employing a *closeness relation* among possible worlds (the atoms of the algebra), often interpreted as a measure of similarity. This closeness relation imposes a certain structure

on the underlying set of possible worlds that is employed within the context of imaging updating methods to determine which possible worlds the probability mass should be redistributed to, typically the closest (most similar) ones to a given one.

Such a closeness relation can also be used to specify the truth conditions of a wide range of conditional connectives, such as variably strict conditionals [19] and preferential conditionals [3, 21, 11]. The informal idea is that a conditional $a \rhd b$ is true at a world w if and only if b is true in the closest a-worlds to w, i.e., those closest worlds that make a true. The structure of closeness has been represented through various model-theoretic frameworks, among which ordered models and sphere models are particularly noteworthy. Sphere models were prominently used by Lewis [19] as the intended semantic device for specifying the truth conditions of variably strict conditionals (including counterfactual conditionals), while ordered models have been used, for example, to specify the truth conditions of preferential conditionals. Consequently, this closeness relation of possible worlds has been employed both in specifying imaging-like updating methods and in establishing a semantics for a large class of conditional connectives. Hence, intuitively, one might expect a deep relationship between imaging-like updating methods and conditionals based on a closeness relation, given that both rely on the same structure.

Our investigation will specifically focus on these two dimensions: imaging-like updating methods and conditionals, both having a semantics grounded on a closeness relation. Given our aim for generality, we will slightly deviate from the standard model-theoretic approach to conditionals and updating methods that relies on sphere models or ordered models to represent closeness relations, and we will utilize models based on *selection functions*. Notably, certain classes of ordered models and sphere models have equivalent representations in terms of selection functions (see, e.g., [18] and also [13]). One of the main advantages of the selection function is its technical utility: it is a very general tool for representing a general structure over worlds. However, one of its main shortcomings is conceptual: while orders and spheres offer an intuitive representation of a certain closeness relation, the relation induced by a selection function can sometimes be obscure and abstract.

In the present contribution we aim to provide a further generalization of imaging methods by relying on this more general setting of selection functions. Simultaneously, we aim to present results that conceptually enhance the understanding of certain updating procedures and selection function models. Inspired by this generality, we will primarily work within an algebraic setting, which also generalizes possible worlds models and imaging updating methods. Our basic framework will consist of a finite Boolean algebra **A**, whose domain is denoted by the corresponding non-bold italic letter A, and whose atoms, forming the set at(**A**), will be thought as possible worlds. The finite algebra **A** will be equipped with a probability function P, whence at(**A**) will be endowed with a probability distribution that we still denote by P. The structure on A that we need to establish our general updating methods is represented by two basic tools (more technical details will be provided in the next sections):

- a selection function $f : A \times \text{at}(\mathbf{A}) \to A$ that, for every atom $\alpha \in \text{at}(\mathbf{A})$ and every $a \in A$, specifies an element of A that correspond to the set of worlds closest to α according to a;

- a function $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$ that will be used to determine how to distribute the probability mass of each $\alpha \in \text{at}(\mathbf{A})$ to its closest worlds $f(a, \alpha)$.

While the next section will introduce the basic notions and results that constitute the ground of our investigation, Section 3 will connect selection function-based conditionals with modalities and probabilities, while Section 4 will more directly deal with general updating methods. We will end the present paper with some conclusions that will be put forward in Section 5.

2   Selection Functions and Conditionals

Let us begin by establishing the fundamental components that define our framework. Boolean algebras will be presented in their usual language $(\wedge, \vee, \neg, \bot, \top)$ of type $(2,2,1,0,0)$ and where the implication operator $\rightarrow$ will be defined as $a \rightarrow b = \neg a \vee b$. All Boolean algebras that we consider in this paper are assumed to be finite; a generic Boolean algebra will be denoted by $\mathbf{A}$ and the set of its atoms is given by $\mathrm{at}(\mathbf{A}) = \{\alpha_1, \ldots, \alpha_n\}$. Let $\wp(X)$ stand for the powerset of a set X. Thus, $\mathbf{A}$ and $\wp(\mathrm{at}(\mathbf{A}))$ are isomorphic algebras whence we shall henceforth identify elements of A with subsets of $\mathrm{at}(\mathbf{A})$ and use a typical set-theoretical notation without danger of confusion. For instance, for $a \in A$, we will write $|a|$ to denote the cardinality of the subset of $\mathrm{at}(\mathbf{A})$ below a itself. Moreover, for $a, b \in A$ we write $a \subseteq b$ to state that a lies below b in the lattice order of $\mathbf{A}$ while, when $\alpha \in \mathrm{at}(\mathbf{A})$ and $a \in A$ we will simply write $\alpha \in a$ to denote that α is an atom among those that lie below a.

A selection function for $\mathbf{A}$ is a mapping $f : A \times \mathrm{at}(\mathbf{A}) \rightarrow A$ that associate to each pair of an atom and an element of A another element of A that we will identified with a set of atoms of $\mathbf{A}$ as observed above. Initially, we do not impose any further properties on this function. Intuitively, the selection function specifies the set of closest worlds (atoms) to a given world, relative to a certain element. For instance, $f(a, \alpha)$ represents the set of worlds closest to α, as determined by a. For every selection function $f : A \times \mathrm{at}(\mathbf{A}) \rightarrow A$, let us define an *selection function-based conditional* operator (or simply a *conditional* operator)

$$\rhd_f : A \times A \rightarrow A$$

as follows: for all $a, b \in A$,

$$a \rhd_f b = \{\alpha \in \mathrm{at}(\mathbf{A}) \mid f(a, \alpha) \subseteq b\}. \tag{1}$$

In other words, for all $\alpha \in \mathrm{at}(\mathbf{A})$,

$$\alpha \in a \rhd_f b \text{ iff } f(a, \alpha) \subseteq b.$$

As we have already anticipated, this algebraic setting has a very intuitive possible-worlds semantic counterpart: an atom α belonging to $a \rhd_f b$ can be regarded as the possible world α making $a \rhd_f b$ true. Consequently, given an intuitive interpretation of the selection function in terms of a closeness relation, the above condition is telling us that $a \rhd_f b$ is true at α if and only if all the closest a-worlds make b true.

This closeness relation, specified by f, is at this stage very general while it can be instantiated in different ways to represent, for instance, an order of similarity [19] or normality among worlds [3]. The logical properties of $\rhd_f$ will be induced by the constraints imposed on f, i.e., the type of closeness relation we aim to represent. Indeed, having imposed no specific properties on the selection function f does not inherently explain why the operator $\rhd_f$ defined as above should be considered a *conditional* operator. For example, it might be desirable that $a \rhd_f a$ is always true, or that $\rhd_f$ satisfies the modus ponens inference rule, and so on. All these properties can be enforced by imposing relevant constraints on the closeness relation involved in evaluating the conditional, namely the selection function. Below, we list several relevant constraints, including some that are discussed in [18]: consider a finite Boolean algebra $\mathbf{A}$ and a selection function $f : A \times \mathrm{at}(\mathbf{A}) \rightarrow A$, we can define the following properties for all $\alpha \in \mathrm{at}(\mathbf{A})$:

- $f(\bot, \alpha) = \bot$ (emptyness)
- for $a \in A \setminus \{\bot\}$, $f(a, \alpha) \neq \bot$ (normality)
- for $a \in A$, $f(a, \alpha) \subseteq a$ (identity)
- if $\alpha \in a$, then $\alpha \in f(a, \alpha)$ (centering-1)

- if $\alpha \in a$, then $f(a,\alpha) \subseteq \alpha$ (centering-2)
- if $\alpha \in a$, $f(a,\alpha) = \alpha$ (centering)
- $|f(a,\alpha)| \leq 1$ (uniqueness)
- if $f(a,\alpha) \subseteq b$ and $f(b,\alpha) \subseteq a$, then $f(a,\alpha) = f(b,\alpha)$ (well-order)
- $f(a \vee b, \alpha) \subseteq a$ or $f(a \vee b, \alpha) \subseteq b$ or $f(a \vee b, \alpha) = f(a,\alpha) \vee f(b,\alpha)$ (nesting)

The connection to the corresponding conditional connective is straightforward, and this allows to prove the following result, which readily follows from the conditions imposed on the selection function and equation (1) and from well-known results proved in [18]:

Fact 1. *Consider a finite Boolean algebra* **A** *and a selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$*; the following hold for all* $a,b,c \in A$*, and all* $d \in A \setminus \{\bot\}$*:*

$$\perp \rhd_f a = \top \qquad\qquad\qquad\qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies emptyness}$$
$$d \rhd_f b \subseteq \neg(d \rhd_f \neg b) \qquad\qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies normality}$$
$$a \rhd_f a = \top \qquad\qquad\qquad\qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies identity}$$
$$a \rhd_f b \subseteq \neg a \vee b \qquad\qquad\qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies centering-1}$$
$$a \wedge b \subseteq a \rhd_f b \qquad\qquad\qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies centering-2}$$
$$a \wedge b \subseteq a \rhd_f b \subseteq \neg a \vee b \qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies centering}$$
$$(a \rhd_f \neg b) \vee (a \rhd_f b) = \top \qquad\qquad\qquad \Leftrightarrow \quad f \text{ satisfies uniqueness}$$
$$(a \rhd_f b) \wedge (b \rhd_f a) \subseteq (a \rhd_f c) \leftrightarrow (b \rhd_f c) \qquad \Leftrightarrow \quad f \text{ satisfies well-order}$$
$$((a \vee b) \rhd_f a) \vee ((a \vee b) \rhd_f b) \vee (((a \vee b) \rhd_f c) = ((a \rhd_f c) \wedge (b \rhd_f c))) \quad \Leftrightarrow \quad f \text{ satisfies nesting}$$

For convenience, we may refer to $\rhd_f$ as to a *counterfactual conditional* when f satisfies the following properties: (i) identity, (ii) well-ordering, (iii) nesting, and (iv) centering. We call $\rhd_f$ a *variably strict conditional* when f satisfies (i) identity, (ii) well-ordering, and (iii) nesting. Finally, we use the term *Stalnaker conditional* for $\rhd_f$ when f satisfies (i) identity, (ii) well-ordering, (iii) nesting, (iv) centering, and (v) uniqueness. This terminology is justified by well-established results in the literature implying that that when $\rhd_f$ is a counterfactual conditional, it satisfies precisely the logical principles of Lewis's counterfactuals, namely it obeys the conditional logic C1 (see [18]). When $\rhd_f$ is a variably strict conditional, it obeys the conditional logic C0 (see [18]), which is the basic variably strict conditional logic (see [19]). And when $\rhd_f$ is a Stalnaker conditional, it obeys the logic C2 (see [18]), which corresponds to Stalnaker's logic of conditionals [25].

3   Modalities and Probabilities

In this section, we will establish some key results concerning the relationship between normal modal operators and selection functions. These results will subsequently be employed as a tool to begin exploring the connection between selection function-based conditionals and probability.

Consider a finite Boolean algebra **A** and a selection function $f : A \times \mathrm{at}(\mathbf{A}) \to A$; let the binary relation $R_a^f \subseteq \mathrm{at}(\mathbf{A}) \times \mathrm{at}(\mathbf{A})$ be defined as follows: for all $\alpha \in A$, $R_a^f(\alpha) = f(a,\alpha)$. In other words,

$$\alpha R_a^f \beta \text{ iff } \beta \in f(a,\alpha).$$

For each such R_a^f, let $\square_a^f : A \to A$ be defined as standard normal modal operator: for all $b \in A$

$$\square_a^f(b) = \{\alpha \in \mathrm{at}(\mathbf{A}) \mid R_a^f(\alpha) \subseteq b\} = \{\alpha \in \mathrm{at}(\mathbf{A}) \mid f(a,\alpha) \subseteq b\}.$$

Notice that $\square_a^f$ is, indeed, a normal modal operator in the sense of [2, Definition 1.42], that is to say, $\square_a^f \top = \top$ and $\square_a^f(b \to c) \subseteq (\square_a^f b \to \square_a^f c)$. Moreover, since $R_a^f(\alpha) = f(a, \alpha)$, comparing the latter with (1) we can immediately show the following result:

Fact 2. *For a finite Boolean algebra* **A** *and a selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$; *for all* $a, b \in A$, *it holds that*

$$\alpha \in a \rhd_f b \text{ iff } \alpha \in \square_a^f(b).$$

In other words, for every $a \in A$, $a \rhd_f (\cdot)$ and $\square_a^f(\cdot)$ are the same unary operator on A. This implies that selection function-based conditionals can also be equivalently defined in terms of a family of indexed normal modal operators, $\{\square_a^f \mid a \in A\}$, each of which behaves as a conditional with a fixed antecedent. Consequently, the structure $(\mathbf{A}, \{\square_a^f\}_{a \in A})$ forms a *Boolean algebra with operators* in the sense of [2]. Following standard practice in modal logic, we define $\lozenge_a^f x$ as $\neg \square_a^f \neg x$ and for all $b \in A$

$$\lozenge_a^f(b) = \{\beta \in \mathrm{at}(\mathbf{A}) \mid R_a^f(\beta) \wedge b \neq \bot\}.$$

Treating a conditional (binary) operator as a family of indexed normal modal (unary) operators provides a valuable framework for gaining new insights into the properties of conditional operators. This approach allows us to leverage the well-established behavior of normal modal operators and to demonstrate relevant results that connect the properties of a selection function f (and consequently, the corresponding conditional $\rhd_f$) with the properties of the induced indexed modal operators. The following result contributes to this direction:

Fact 3. *For a finite Boolean algebra* **A** *and selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$ *the following holds for all* $a \in A$:

$$|f(a, \alpha)| = 1 \Leftrightarrow (\text{for all } x \in A, \square_a^f x = \lozenge_a^f x).$$

Proof. The identity $\square x = \lozenge x$ holds in a Boolean algebra with a modal operator $(\mathbf{A}, \square)$ iff its associated Kriple frame $(\mathrm{at}(\mathbf{A}), R)$ satisfies that for all $\alpha \in \mathrm{at}(\mathbf{A})$, there exists a unique $\beta \in \mathrm{at}(\mathbf{A})$ such that $\alpha R \beta$, and hence R is a function from $\mathrm{at}(\mathbf{A})$ to itself, [2]. Thus in particular, $\square_a^f x = \lozenge_a^f x$ for all $x \in A$ if and only if R_a^f is a function and hence, if and only if, for all α, there is a unique β such that $\alpha R_a^f \beta$, that is to say, there exists a unique $\beta \in f(a, \alpha)$. Therefore, if and only if $|f(a, \alpha)| = 1$. $\square$

We now have the basic ingredients to begin investigating the relationship between conditionals and probability using the introduced general framework. For our investigation we will henceforth always assume that probability functions are positive, *i.e.*, they assign 0 only to the impossible event. Given the finite algebraic environment we are adopting, this choice comes at no cost. So, let $P : \mathbf{A} \to [0, 1]$ be a positive probability on a finite Boolean algebra **A** and consider a selection function $f : A \times \mathrm{at}(\mathbf{A}) \to A$. Then, given the above assumptions and definitions, we define

$$P(a \rhd_f b) = \sum_{\alpha \in a \rhd_f b} P(\alpha).$$

That is, the probability of a conditional $a \rhd_f b$ is given by the sum of the probabilities of the atomic states (worlds) where the conditional is satisfied. In other words, $P(a \rhd_f b)$ represents the probability that the conditional $a \rhd_f b$ is true. Directly from Fact 2, we obtain the following.

Fact 4. *For a finite Boolean algebra* **A***, a selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$, *and probability* $P : \mathbf{A} \to [0, 1]$ *the following holds for all* $a, b \in A$:

$$P(a \rhd_f b) = P(\square_a^f(b)).$$

As a direct consequence of the above Fact 4 and by the main results of [17], the map defined as

$$P(a \rhd_f (\cdot)) = P(\square_a^f(\cdot)) \tag{2}$$

is a *belief function* in the sense of Dempster-Shafer theory [5]. Clearly, this observation does not preclude the possibility that $P(a \rhd_f (\cdot))$ is a probability function for some $a \in A$.

Let us now examine in more detail this connection between conditionals and belief functions. First, recall that belief functions differ from probability functions in their treatment of disjunction: probabilities are additive, satisfying $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$, whereas belief functions, denoted *Bel*, are superadditive, that is to say, $Bel(a \vee b) \geq Bel(a) + Bel(b) - Bel(a \wedge b)$.

Besides the axiomatic definition of belief functions, that the interest reader can find for instance in [16] and on which we will not rely for the rest of the present paper, it is convenient to recall how belief functions can be described in terms of their *mass distribution*. Specifically, every belief function is induced by a mass distribution m over a Boolean algebra $\mathbf{A}$, i.e., a function $m : A \to [0,1]$ such that $\sum_{a \in A} m(a) = 1$. From this mass distribution, the belief function is then defined as follows:

$$Bel(a) = \sum_{b \subseteq a} m(a). \tag{3}$$

Observe that, while probability functions on $\mathbf{A}$ are fully determined by distributions on $\mathrm{at}(\mathbf{A})$, belief functions on $\mathbf{A}$ are fully described by mass functions on A itself, and not only on its atoms. This easy observation marks the key difference between the two uncertain measures.

Given this definition, we can more vividly represent the belief function induced by a conditional $\rhd_f$ according to equation 2 by appealing to a generalization of the *imaging-like updating method* for belief functions introduced in [6]. Specifically, let us consider a finite Boolean algebra $\mathbf{A}$, a selection function $f : A \times \mathrm{at}(\mathbf{A}) \to A$, and a probability $P : A \to [0,1]$. For a given element $a \in A$ we can define a mass distribution $m_a : A \to [0,1]$ as follows: $m_a(b) = \sum_{\alpha:f(a,\alpha)=b} P(\alpha)$. Then the corresponding *imaged* belief function is:

$$Bel_a(b) = \sum_{c \subseteq b} m_a(c) = \sum_{c \subseteq b} \left(\sum_{\alpha:f(a,\alpha)=c} P(\alpha) \right) = \sum_{\alpha:f(a,\alpha) \subseteq b} P(\alpha). \tag{4}$$

As a consequence, we obtain the following:

Fact 5. *Consider a finite Boolean algebra* $\mathbf{A}$*, a selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$*, and a probability* $P : \mathbf{A} \to [0,1]$*, the following holds for all* $a, b \in A$*:*

$$P(a \rhd_f (\cdot)) = P(\square_a^f(\cdot)) = Bel_a(\cdot)$$

$$P(a \rhd_f b) = P(\square_a^f(b)) = Bel_a(b)$$

Specifically, the probability of a selection function conditional $a \rhd_f b$ is equal to the corresponding belief function of the consequent imaged on the antecedent, i.e. $Bel_a(b)$.

This general result implies that the probability of a broad class of conditionals can be characterized in terms of an updated belief function. For instance, variably strict conditionals, which encompass both Stalnaker conditionals and Lewis counterfactuals (see [19, 18]), are known to be characterizable as selection function-based conditionals. Similarly, preferential conditionals [21, 11, 3] also fall under this category. Indeed, the semantics of preferential conditionals is typically defined using preorders, such that a preferential conditional $a \rhd_f b$ is true at a world α if and only if the minimal a-worlds with respect to the preorder associated with α are also b-worlds. This semantics can be reformulated such that the

set of minimal a-worlds, with respect to the preorder associated with α, constitutes a selection function $f(a, \alpha)$. Hence, the aforementioned result also holds for preferential conditionals. This result generalizes Lewis's [20] work on Stalnaker conditionals. Lewis demonstrated that the probability of a Stalnaker conditional $a \rhd_f b$ corresponds to the probability of the consequent b imaged on the antecedent a. Essentially, the probability of a Stalnaker conditional aligns with a specific updated probability distribution. Given that belief functions generalize probabilities, our finding indicates that the probability of selection function-based conditionals generally follows an updated belief function, and in specific instances, for instance in the case of Stalnaker conditionals, this updated belief function can indeed reduce to a probability, thereby implying Lewis's original result (further details are in subsequent sections).

Therefore, it is legitimate to ask under what conditions the belief function induced by a conditional is also a probability function. The next result provides an answer to this question, bringing together some of the results seen so far and characterizing the conditions under which the map defined in (2) is a probability function.

Proposition 1. *Consider a finite Boolean algebra* **A**, *a selection function* $f : A \times \mathrm{at}(\mathbf{A}) \to A$, *and a positive probability* $P : A \to [0, 1]$, *for all* $a \in A$, *the following conditions are equivalent:*

(i) $P(a \rhd_f (\cdot)) : A \to [0, 1]$ is a probability function,

(ii) $|f(a, \alpha)| = 1$,

(iii) R_a^f is a function,

(iv) for all $x \in A$, $\square_a^f x = \Diamond_a^f x$.

Proof. The equivalence between (ii), (iii), and (iv) follows from Fact 3 and the definition of R_a^f. We now prove that (iv) implies (i). Indeed, if $\square_a^f x = \Diamond_a^f x$ holds for all $x \in A$, we have that $\square_a^f$ distributes over disjunction, namely $\square_a^f (c \vee b) = \square_a^f c \vee \square_a^f b$, and so, by Fact 5, we would have that $Bel_a(c \vee b) = Bel_a(b) + Bel(b) - Bel_a(c \wedge b)$. Namely, the belief function Bel_a is additive and hence a probability; namely $P(a \rhd_f (\cdot)) : A \to [0, 1]$ is a probability function. We now prove that (i) implies (ii). Let us reason by contraposition and assume that f doesn't satisfy uniqueness, namely that there is an atom α and an element $a \in A$ such that $|f(a, \alpha)| > 1$ or $f(a, \alpha) = \bot$. If the former holds, let $b = f(a, \alpha)$, then, by definition of Bel_a, we would have that $Bel_a(b) > 0$, since $m_a(b) > 0$. Thus, take $\beta \in b$ so that $b = \beta \vee (b \wedge \neg\beta)$, and then one has $Bel_a(b) > Bel_a(\beta) + Bel_a(b \wedge \neg\beta)$, whence Bel_a is not additive. Finally, if $f(a, \alpha) = \bot$, then $Bel_a(\bot) > 0$. Therefore, Bel_a is not a probability either. So, by Fact 5, $P(a \rhd_f (\cdot))$ is not a probability. $\square$

The above results are general findings concerning conditionals and probability, and they can be interpreted as having limitative implications. On one hand, they demonstrate that the probability of a selection function-based conditional aligns with imaged belief functions. On the other hand, they establish that the belief function induced by a conditional is a probability function if and only if the selection function underlying the conditional satisfies the uniqueness property. Given these general results, we can now proceed to investigate whether more specific relationships exist between updating methods for probability and selection function-based conditionals.

4   Dealing with General Updates

As briefly mentioned in the introduction, Lewis [20] introduced *imaging* as an alternative to conditionalization for updating a probability measure $P : \mathbf{A} \to [0, 1]$. Roughly, this procedure changes the initial

probability P by acting as if a certain event a holds. This update is accomplished by leveraging a similarity structure among possible worlds (or, equivalently, a selection function): the updated a-imaged probability P_a is obtained by shifting the probability mass of each atom α to its closest a-world. Crucially, this updating procedure relies on a uniqueness assumption: for each $\alpha \in \text{at}(\mathbf{A})$, there is only one closest a-world to which α transfers its probability mass $P(\alpha)$. Gärdenfors [12] further generalized Lewis's imaging by relaxing this uniqueness assumption, allowing for multiple closest worlds. Building on this prior work and drawing inspiration from Güenther [14], we now introduce a general framework for defining updating methods for probabilities that generalizes both Lewis and Gärdenfors's imaging.

First, given a finite Boolean algebra $\mathbf{A}$ and a selection function $f : A \times \text{at}(\mathbf{A}) \to A$, we define a *distribution function for f* as a map $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$ satisfying the following constraints:

- $\lambda(a,\alpha)(\beta) = 0$ if and only if $\beta \notin f(a,\alpha)$

- $\sum_{\beta \in f(a,\alpha)} \lambda(a,\alpha)(\beta) = 1$

Observe that the second item above states that, for all $a \in A$ and $\alpha \in \text{at}(\mathbf{A})$, $\lambda(a,\alpha)$ is a probability distribution on $f(a,\alpha)$.

Now, consider a finite Boolean algebra $\mathbf{A}$, a positive probability $P : \mathbf{A} \to [0,1]$, a selection function $f : A \times \text{at}(\mathbf{A}) \to A$, and distribution function $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$ for f. Given an element $a \in A \setminus \{\bot\}$, we can define a probability distribution P_a^λ over $\text{at}(\mathbf{A})$ as follows: for all $\beta \in \text{at}(\mathbf{A})$

$$P_a^\lambda(\beta) = \sum_{\alpha : \beta \in f(a,\alpha)} \lambda(a,\alpha)(\beta) \cdot P(\alpha) \tag{5}$$

Specifically, the probability distribution $P_a^\lambda(\cdot)$ is obtained by redistributing the original probability mass of each atom α among all atoms in $f(a,\alpha)$, resulting in a new probability distribution. Hence, the corresponding updated probability function can be expressed as:

$$P_a^\lambda(b) = \sum_{\beta \in b} P_a^\lambda(\beta) = \sum_{\beta \in b} \left(\sum_{\alpha : \beta \in f(a,\alpha)} \lambda(a,\alpha)(\beta) \cdot P(\alpha) \right). \tag{6}$$

$P_a^\lambda(\cdot)$ represents an updated probability distribution where the shift of probability mass is performed according to the selection function f. The constraints imposed on the distribution function λ ensure that $P_a^\lambda(\cdot)$ is indeed a probability distribution. Moreover, all the probability mass is transferred to the atoms selected by f. Indeed, the following result can be proven from equation 6 and the definition of the distribution function λ:

Fact 6. *Consider a finite Boolean algebra $\mathbf{A}$, a selection function $f : A \times \text{at}(\mathbf{A}) \to A$ and a distribution mass $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$. For $a \in A \setminus \{\bot\}$, $P_a^\lambda(\cdot)$ is a probability function*

According to the general updating method for probabilities defined above, which utilizes the distribution function λ, several different updating procedures can be represented, including conditionalization [14], imaging [20], and general imaging [12]. Furthermore, since λ allows for considerable flexibility in the redistribution of probability mass, numerous other updating procedures, beyond these standard and common ones, can also be defined within this framework.

Having established this foundation, an important question remains: what is the relationship between selection function-based conditionals and λ-updating methods for probability? In light of the preceding results, we can demonstrate that these updating methods can be characterized as a probability induced

by a conditional only in specific cases, thus proving that selection function-based conditionals cannot capture the entire scope of λ-updating procedures.

First, let us prove a general result concerning the relationship between the probability of conditionals and updating methods, namely that the probability of a conditional $a \rhd_f b$ (with f satisfying normality) will always be less than or equal than the corresponding λ-updated probability of b given a:

Fact 7. *Consider a finite Boolean algebra* $\mathbf{A}$, *a selection function* $f : A \times \text{at}(\mathbf{A}) \to A$ *satisfying normality, and a distribution mass* $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$. *For* $a \in A \setminus \{\bot\}$, *and for all* $b \in A$,

$$P(a \rhd_f b) \leq P_a^\lambda(b)$$

Proof. (Sketch) We show that for any $\alpha \in a \rhd_f b$, the initial probability mass of $P(\alpha)$ contributes to the calculation of $P_a^\lambda(b)$. Specifically, let us consider $\alpha \in \text{at}(\mathbf{A})$ such that $\alpha \in a \rhd_f b$. This implies that $f(a,\alpha) \subseteq b$. According to the definition of $P_a^\lambda(\cdot)$, the probability mass $P(\alpha)$ is redistributed among the elements in $f(a,\alpha)$. However, since $f(a,\alpha) \subseteq b$, we know that for every $\beta \in f(a,\alpha)$, $\beta \in b$. This implies that the mass $P(\alpha)$ from every $\alpha \in a \rhd_f b$ is redistributed among atoms that are below b. Consequently, this mass will be included in the summation that yields $P_a^\lambda(b)$. Thus, we have $\sum_{\alpha \in a \rhd_f b} P(\alpha) \leq P_a^\lambda(b)$, which is equivalent to $P(a \rhd_f b) \leq P_a^\lambda(b)$. $\qquad\square$

We now have all the necessary components to show the conditions under which the probability of a conditional corresponds to an updated probability.

Theorem 1. *Consider a finite Boolean algebra* $\mathbf{A}$, *a selection function* $f : A \times \text{at}(\mathbf{A}) \to A$ *satisfying normality, a distribution function* $\lambda : A \times \text{at}(\mathbf{A}) \to [0,1]^{|\text{at}(\mathbf{A})|}$, *and a positive probability* $P : A \to [0,1]$. *The following holds:*

$$(\text{for all } a \in A \setminus \{\bot\} \text{ and } b \in A, \; P(a \rhd_f b) = P_a^\lambda(b)) \Leftrightarrow f \text{ satisfies uniqueness}$$

Proof. (Sketch) ($\Leftarrow$). This direction is straightforward. ($\Rightarrow$) By contraposition, assume that f doesn't satisfy uniqueness, namely there are α and a such that $|f(a,\alpha)| > 1$, by normality. Then, by Proposition 1, we have that $P(a \rhd_f (\cdot))$ is not a probability function, hence, a fortiori, it cannot be $P_a^\lambda(\cdot)$ which is a probability by Fact 6; therefore there must be some $a \in A \setminus \{\bot\}$ and $b \in A$ such that $P(a \rhd_f b) \neq P_a^\lambda(b)$. $\qquad\square$

In conclusion of this section, let us examine the previous result in more detail. First of all, let us observe that if f satisfies the uniqueness constraint (along with normality), then for all $a \in A \setminus \{\bot\}$, $P(a \rhd_f (\cdot))$ and $P_a^\lambda(\cdot)$ coincide. This is because, under uniqueness, $\lambda(a,\alpha)$, being a probability distribution on $f(a,\alpha) = \{\beta\}$, it can only take values of 0 or 1; specifically, $\lambda(a,\alpha)(\beta) = 1$ and 0 otherwise. This means that each world (or atom) α transfers its entire probability mass to its unique closest a-world, as determined by f. Hence, due to uniqueness, the mass m_a, reduces to a probability distribution.

Conversely, when f does not satisfy uniqueness, it generally holds that for some $a,b \in A$, with $a \neq \bot$, $P(a \rhd_f b) < P_a^\lambda(b)$ as the following example shows.

Example 1. *Let us consider the algebra* $\mathbf{A}$ *of three atoms, i.e.* $\text{at}(\mathbf{A}) = \{\alpha_1, \alpha_2, \alpha_3\}$. *Furthermore, conforming to our convention, let us represent the elements of* A *as subsets of* $\text{at}(\mathbf{A})$.

Let us consider the selection function f *such that* $f(\{\alpha_2, \alpha_3\}, \alpha_1) = \{\alpha_2, \alpha_3\}$, $f(\{\alpha_2, \alpha_3\}, \alpha_2) = \{\alpha_2\}$ *and* $f(\{\alpha_2, \alpha_3\}, \alpha_3) = \{\alpha_3\}$. *Notice that* f *does satisfy normality, yet not the uniqueness property because* $|f(\{\alpha_2, \alpha_3\}, \alpha_1)| = 2$.

Let P *be a positive probability on* $\mathbf{A}$ *and* λ *be a distribution function. Recall from what we stated at the very beginning of this section that for all* $a \in A$ *and* $\alpha \in \text{at}\mathbf{A}$, $\lambda(a,\alpha)$ *is a probability distribution on* $f(a,\alpha)$. *Thus, assuming* λ *to be positive, we have that*

- $\lambda(\{\alpha_2,\alpha_3\},\alpha_1)$ *distributes postively on* $f(\{\alpha_2,\alpha_3\},\alpha_1) = \{\alpha_2,\alpha_3\}$ *and therefore* $\lambda(\{\alpha_2,\alpha_3\},\alpha_1)(\alpha_2) > 0$ *and* $\lambda(\{\alpha_2,\alpha_3\},\alpha_1)(\alpha_3) > 0$;

- $\lambda(\{\alpha_2,\alpha_3\},\alpha_2)$ *can only assign* 1 *to* α_2 *being* α_2 *the unique element in its domain.*

Then, by definition:

- $P(\{\alpha_2,\alpha_3\} \rhd_f \alpha_2) = P(f(\{\alpha_2,\alpha_3\},\alpha_2)) = P(\alpha_2)$;

- $P^{\lambda}_{\{\alpha_2,\alpha_3\}}(\alpha_2) = \lambda(\{\alpha_2,\alpha_3\},\alpha_1)(\alpha_2) \cdot P(\alpha_1) + \lambda(\{\alpha_2,\alpha_3\},\alpha_2)(\alpha_2) \cdot P(\alpha_2) = \lambda(\{\alpha_2,\alpha_3\},\alpha_1)(\alpha_2) \cdot P(\alpha_1) + 1 \cdot P(\alpha_2)$.

Therefore, since P is positive and $\lambda(\{\alpha_2,\alpha_3\},\alpha_1)(\alpha_2) > 0$,

$$P(\{\alpha_2,\alpha_3\} \rhd_f \alpha_2) < P^{\lambda}_{\{\alpha_2,\alpha_3\}}(\alpha_2).$$

5   Conclusion

The present work aimed to demonstrate that the probability of selection function conditionals can be characterized in terms of updated belief functions within the framework of Dempster-Shafer theory. Conversely, classical Bayesian probabilistic updates (defined in terms of λ) cannot be characterized as the probability of a selection function conditional, except in a highly restricted case, specifically when the set of closest worlds involved in defining both the update and the conditional is a singleton. These results are quite general, as selection function conditionals represent a broad class of conditionals, including variably strict conditionals and preferential conditionals.

The results presented here can be further generalized to gain new insights into the logic of conditionals. In particular, the finding that the probability of selection function-based conditionals can be represented as a belief function suggests that this type of conditional might be representable in terms of a standard normal modal operator from modal logic. Indeed, our results in Section 3 can be further generalized to explore whether selection function-based conditionals can be represented as $a \rhd_f b \equiv \Box(a \to b)$, where $\to$ can be another binary connective. Previous work in this direction by [22] has shown that Lewis counterfactuals $a \rhd_f b$ can be represented as $\Box(b \mid a)$ within the framework of Boolean algebras of conditionals [8], where $(\cdot \mid \cdot)$ is a conditional that satisfies suitable algebraic counterparts of the laws of conditional probability. Given that Lewis counterfactuals are a specific type of selection function conditional, our findings in this work suggest that the representation result in [22] can also be extended to other kinds of selection function conditionals by imposing appropriate constraints on $\Box$.

On a more conceptual level, we have shown that a broad class of updating methods, including conditionalization, cannot be interpreted as the probability that a certain conditional connective holds. A question remains open regarding whether there exists a specific operator whose probability can represent these kinds of updating methods. Our results here indicate that the possible candidates are limited.

References

[1] Ernest W. Adams (1975): *The Logic of Conditionals*. Reidel, Dordrecht, doi:10.1007/978-94-015-7622-2.

[2] Patrick Blackburn, Marteen de Rijke, Yde Venema (2002): *Modal Logic*. Cambridge University Press, Cambridge, doi:10.1017/CBO9781107050884.

[3] John P. Burgess (1981): *Quick Completeness Proofs for Some Logics of Conditionals*. Notre Dame Journal of Formal Logic 22(1), pp. 76–84, doi:10.1305/ndjfl/1093883341.

[4] Brian F. Chellas (1975): *Basic conditional logic*. Journal of Philosophical Logic 4(2), pp. 133–153, doi:10.1007/bf00693270.

[5] Arthur P. Dempster (1967): *Upper and lower probabilities induced by a multivalued mapping*. Ann. Math. Stat 38(2), pp. 325–339, doi:10.1214/aoms/1177698950.

[6] Didier Dubois & Henri Prade (1994): *A survey of belief revision and updating rules in various uncertainty models*. International Journal of Intelligent Systems 9(1), pp. 61–100, doi:10.1002/int.4550090105.

[7] Paul Égré, Lorenzo Rossi & Jan Sprenger (forthcoming): *Certain and Uncertain Inference with Indicative Conditionals*. Australasian Journal of Philosophy.

[8] Tommaso Flaminio, Lluis Godo & Hykel Hosni (2020): *Boolean algebras of conditionals, probability and logic*. Artificial Intelligence 286, p. 103347, doi:10.1016/j.artint.2020.103347.

[9] Tommaso Flaminio, Giuliano Rosella & Lluis Subirana (2025): *From conditionalization to imaging through the triviality glass*. Submitted Manuscript.

[10] Bas C. van Fraassen (1976): *Probabilities of Conditionals*. In: Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Springer Netherlands, pp. 261–308, doi:10.1007/978-94-010-1853-1_10.

[11] Marianna Girlando, Sara Negri & Nicola Olivetti (2021): *Uniform labelled calculi for preferential conditional logics based on neighbourhood semantics*. Journal of Logic and Computation 31(3), pp. 947–997, doi:10.1093/logcom/exab019.

[12] Peter Gärdenfors (1982): *Imaging and Conditionalization*. Journal of Philosophy 79, pp. 747–760, doi:10.2307/2026039.

[13] Gösta Grahne (1998): *Updates and Counterfactuals*. Journal of Logic and Computation 8, pp. 87–117, doi:10.1093/logcom/8.1.87.

[14] Mario Günther (2022): *Probabilities of Conditionals and Imaged Probabilities*. Draft manuscript.

[15] Alan Hájek (1989): *Probabilities of conditionals - Revisited*. Journal of Philosophical Logic 18(4), doi:10.1007/bf00262944.

[16] Joseph Y. Halpern (2003): *Reasoning about Uncertainty*. The MIT press, Cambridge, Massachusetts London, England, doi:10.7551/mitpress/10951.001.0001.

[17] David Harmanec, George J. Klir & Germano Resconi (1994): *On modal logic interpretation of Dempster-Shafer theory of evidence*. International Journal of Intelligent Systems 9(10), pp. 941–951, doi:10.1002/int.4550091003.

[18] David Lewis (1971): *Completeness and decidability of three logics of counterfactual conditionals 1*. Theoria 37(1), pp. 74–85, doi:10.1111/j.1755-2567.1971.tb00061.x.

[19] David Lewis (1973): *Counterfactuals*. Blackwell, Oxford, doi:10.2307/2273738.

[20] David Lewis (1976): *Probabilities of Conditionals and Conditional Probabilities*. Philosophical Review 85(3), pp. 297–315, doi:10.2307/2184045.

[21] Sara Negri & Nicola Olivetti (2015): *A Sequent Calculus for Preferential Conditional Logic Based on Neighbourhood Semantics*, pp. 115–134. Springer International Publishing, doi:10.1007/978-3-319-24312-2_9.

[22] Giuliano Rosella, Tommaso Flaminio & Stefano Bonzio (2023): *Counterfactuals as Modal Conditionals, and Their Probability*. Artificial Intelligence 323(C), p. 103970, doi:10.1016/j.artint.2023.103970.

[23] Paolo Santorio (forthcoming): *Probabilities of Counterfactuals are Counterfactual Probabilities*. Journal of *Philosophy*.

[24] Glenn Shafer (1976): *A Mathematical Theory of Evidence*. Princeton University Press, doi:10.2307/1268172.

[25] Robert Stalnaker (1968): *A Theory of Conditionals*. In Nicholas Rescher, editor: *Studies in Logical Theory*, *American Philosophical Quarterly Monograph Series* 2, Blackwell, Oxford, pp. 98–112, doi:10.1007/978-94-009-9117-0_2.

[26] Robert Stalnaker (1970): *Probability and Conditionals*. Philosophy of Science 37, pp. 64–80, doi:10.1086/288280.