Journal of Applied Non-Classical Logics
Vol. 00, No. 00, Month 201X, 1–30

RESEARCH ARTICLE

From KLM-Style Conditionals to Defeasible Modalities, and Back

Katarina Britz[a] and Ivan Varzinczak[b]

[a] *CSIR-SU CAIR, Stellenbosch University, South Africa*
[b] *CRIL, Univ. Artois & CNRS, France*

(*Received 00 Month 201X; final version received 00 Month 201X*)

We investigate an aspect of defeasibility that has somewhat been overlooked by the non-monotonic reasoning community, namely that of defeasible modes of reasoning. These aim to formalise defeasibility of the traditional notion of necessity in modal logic, in particular of its different readings as action, knowledge and others in specific contexts, rather than defeasibility of conditional forms. Building on an extension of the preferential approach to modal logics, we introduce new modal operators with which to formalise the notion of defeasible necessity and distinct possibility, and that can be used to represent expected effects, refutable knowledge, and so on. We show how KLM-style conditionals can smoothly be integrated with our richer language. We also propose a tableau calculus which is sound and complete with respect to our modal preferential semantics, and of which the computational complexity remains in the same class as that of the underlying classical modal logic.

Keywords: Knowledge representation and reasoning; non-monotonic reasoning; modal logic; preferential semantics; defeasible modalities; tableaux

1. Introduction

Accounts of defeasible reasoning, as traditionally studied in the literature on counterfactuals and non-monotonic reasoning, have focused mostly on one aspect of defeasibility (or exceptionality), namely that of *argument forms* or *conditionals*. Such is the case in conditional logics (Lewis, 1973; Stalnaker, 1968) as well as in the approach to non-monotonic reasoning by Kraus et al. (1990) and Lehmann and Magidor (1992), known as the KLM approach, and related frameworks (Boutilier, 1994; Britz, Heidema, & Meyer, 2008, 2009; Britz, Meyer, & Varzinczak, 2011a, 2012; Crocco & Lamarre, 1992; Friedman & Halpern, 2001; Giordano, Gliozzi, Olivetti, & Pozzato, 2009a, 2009b, 2013, 2015). For instance, in the KLM approach, (propositional) defeasible consequence relations $\mid\!\sim$ with a preferential semantics (Lewis, 1974; Shoham, 1988) are studied. In this setting, the meaning of a defeasible statement (or a 'conditional', as it is sometimes referred to) of the form $\alpha \mid\!\sim \beta$ is that "all normal α-worlds are β-worlds", leaving it open for α-worlds that are, in a sense, exceptional not to satisfy β. With the theory that has been developed around this notion it becomes possible to cope with exceptionality when performing reasoning, as in the well-known Tweety example: normally, birds fly; penguins are birds, but normally, penguins do not fly.

There are of course many other appealing and equally useful aspects of defeasibility besides that of arguments. These include notions such as typicality (Booth, Casini, Meyer, & Varzinczak, 2015; Booth, Meyer, & Varzinczak, 2012, 2013; Giordano et al., 2009b), concerned with the most typical cases or situations (or even the most typical representatives of a class), and belief plausibility (Baltag & Smets, 2006, 2008), which relates to the most plausible epistemic possibilities held by an agent, amongst others. It turns out that with KLM-style defeasible statements one cannot capture these aspects of defeasibility. This has to do partly with the syntactic restrictions imposed on $\mid\!\sim$, namely no nesting of conditionals, but, more fundamentally, it relates to where and how the notion of normality is used in such statements. Indeed, in a KLM defeasible statement $\alpha \mid\!\sim \beta$, the normality spotlight is somewhat put on α, as though normality was a property of the premise rather than the conclusion. Whether or not the situations in which β holds are normal, plays no role in the reasoning that is carried out. Moreover, in the original KLM framework, normality is also linked to the premise as a whole, rather than its constituents. Technically, this meant one could not refer directly to normality of a sentence in the scope of logical operators. This limitation is overcome by taking a (modal) conditional approach *à la* Boutilier (1994) or Governatori et al. (2012) — the resulting conditional logics are sufficiently general to allow for the expression of a number of different forms of defeasible reasoning in modal logics. However, the considered modalities are still the classical ones and the emphasis remains on the defeasibility of either conditionals or rules — again, of arguments forms.

In this paper, we investigate a related, but incomparable, notion which we refer to as defeasible *modes of inference* (Britz & Varzinczak, 2012). These amount to defeasible versions of the traditional notion of necessity in modal logics and its different readings as action, knowledge and others in specific application domains. For instance, in an action context, one can say that normally the outcome of a given action a is α. However, we may also want to state that the normal outcome of a is α (Laverny & Lang, 2005), which is different from the former statement. To see why, the first statement says that in the most normal worlds, the result of performing the action a is *always* α, whereas in the second one it is in the most normal situations *resulting* from a's execution that α holds.

For a concrete example, assume one arrives at a dark room and wants to toggle the light switch. Exceptionally, the light will not turn on. This can be either because the light bulb is blown (the current situation is abnormal) or because an overcharge resulted from switching the light (the action behaves abnormally). In the former case, the normality of the situation or state before the action is assessed, whereas in the latter the relative normality of the situation is assessed against all possible outcomes. Here we are interested in the formalisation of the latter type of statement, where it becomes important to shift the notion of normality from the premise of an inference to — in this example — the effect of an action and, importantly, use it in the scope of other logical constructors.

The importance of defeasibility in specific modes of reasoning is also illustrated by the following example. Although one may envisage a situation where the velocity of a sub-atomic particle in a vacuum is greater than c (the speed of light in a vacuum), it is in a sense known that c is the highest possible speed. We are then entitled to derive factual consequences of this scientific theory that also will be 'known'. This venturous version of knowledge, which patently differs from belief, provides for a more fine-grained notion of knowledge that may turn out to be wrong but which is not of the same nature as suppositions or beliefs. (We do not say "we believe the speed of light is the limit".) Our proposal is not aimed at challenging the position of knowledge as indefeasible, justified true belief (Gettier, 1963; Lehrer & Paxson, 1969), but rather provides an extension to epistemic modal logics to allow for reasoning with a notion that we shall refer to as "refutable knowledge".

The remaining of the present text is structured as follows: after setting up the notation and terminology that we shall follow in this paper (Section 2), we define a modal language enriched with defeasible modalities allowing for the formalisation of defeasible versions of modes of inference (Section 3). In particular, we discuss what an appropriate semantics for this new modal language should be by examining some candidates from the literature.

Following that, we revisit Britz et al.'s preferential semantics for modal logic (Britz et al., 2011a, 2012) and argue that it is an adequate semantics for our framework (Section 4). One of the reasons is that it allows for a smooth integration of KLM-style conditionals with our richer language, which we address in Section 5. Following that, we define a tableau system for this broader framework allowing for both defeasible modalities and defeasible conditionals (Section 6), and show that it is sound and complete with respect to our preferential semantics. After a discussion of, and comparison with, related work (Section 7), we conclude with some comments and directions for further investigation.

The present submission is an elaborated extended version of the work read at the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK), which is available on arXiv (http://arxiv.org/abs/1310.6409).

2. Logical Preliminaries

In this work, we shall assume the reader is familiar with modal logic (Blackburn, Benthem, & Wolter, 2006; Chellas, 1980). The purpose of this section is mainly to make explicit the terminology and notation that we shall employ.

We assume a set of *atomic propositions* $\mathcal{P}$, using the logical connectives $\wedge$ (conjunction), $\neg$ (negation), and a set of modal operators $\Box_i$, $1 \leq i \leq n$. Propositions are denoted by $p, q, \ldots$, and sentences by $\alpha, \beta, \ldots$, constructed in the usual way according to the following rule ($1 \leq i \leq n$):

$$\alpha \ ::= p \mid \neg\alpha \mid (\alpha \wedge \alpha) \mid \Box_i\alpha$$

All the other truth-functional connectives ($\vee$, $\rightarrow$, $\leftrightarrow$, ...) are defined in terms of $\neg$ and $\wedge$ in the usual way. Given $\Box_i$, $1 \leq i \leq n$, with $\Diamond_i$ we denote its *dual* modal operator, i.e., for any α, $\Diamond_i\alpha := \neg\Box_i\neg\alpha$. We use $\top$ as an abbreviation for $p \vee \neg p$ and $\bot$ as an abbreviation for $p \wedge \neg p$, for some $p \in \mathcal{P}$.

With $\mathcal{L}^{\Box}$ we denote the *language* of all modal sentences, which is understood as the set of symbol sequences generated according to the rules above. When writing down concepts of $\mathcal{L}^{\Box}$, we shall omit parentheses whenever they are not essential for disambiguation.

The semantics is the standard possible-worlds one:

Definition 1 (Kripke Model). *A Kripke model is a tuple $\mathscr{M} := \langle W, R, V\rangle$ where W is a (non-empty) set of possible worlds, $R := \langle R_1, \ldots, R_n\rangle$, where each $R_i \subseteq W \times W$ is an accessibility relation on W, $1 \leq i \leq n$, and $V : W \longrightarrow \{0,1\}^{\mathcal{P}}$ is a valuation function mapping possible worlds into propositional valuations.*

As an example, Figure 1 depicts the Kripke model $\mathscr{M}_1 = \langle W_1, R_1, V_1\rangle$, where $W_1 := \{w_i \mid 1 \leq i \leq 4\}$, $R_1 := \langle R_a, R_b\rangle$, with $R_a := \{(w_1, w_1), (w_1, w_2), (w_3, w_3), (w_3, w_4)\}$, and $R_b := \{(w_1, w_4), (w_2, w_1), (w_2, w_2), (w_2, w_4), (w_4, w_3)\}$, and V_1 is the obvious valuation function.

In our pictorial representations of models, we shall represent propositional valuations as sequences of 0s and 1s, and with the obvious implicit ordering of atoms. Thus, for the logic generated from p and q, the valuation in which p is true and q is false will be represented as 10. We shall use $w, u, v, \ldots$ (possibly decorated with primes) to denote possible worlds.

Sentences of $\mathcal{L}^{\Box}$ are true or false relative to a possible world in a given Kripke model. This is formalised by the following truth conditions:

Definition 2 (Truth Conditions). *Let $\mathscr{M} = \langle W, R, V\rangle$ and $w \in W$:*

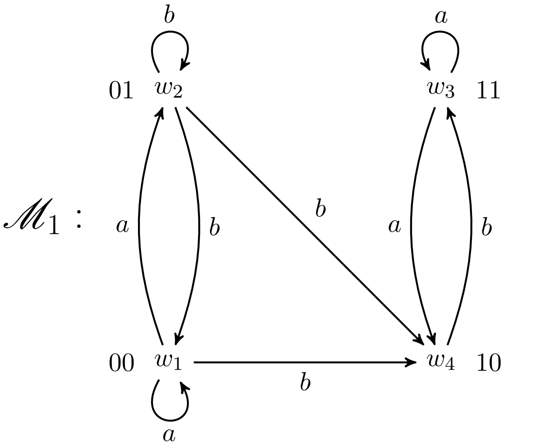


Figure 1. A Kripke model for $\mathcal{P} = \{p, q\}$ and two modalities, namely a and b.

- *$\mathscr{M}, w \Vdash p$ if and only if $V(w)(p) = 1$;*
- *$\mathscr{M}, w \Vdash \neg\alpha$ if and only if $\mathscr{M}, w \nVdash \alpha$;*
- *$\mathscr{M}, w \Vdash \alpha \land \beta$ if and only if $\mathscr{M}, w \Vdash \alpha$ and $\mathscr{M}, w \Vdash \beta$;*
- *$\mathscr{M}, w \Vdash \Box_i\alpha$ if and only if $\mathscr{M}, w' \Vdash \alpha$ for all w' such that $(w, w') \in R_i$.*

Given $\alpha \in \mathcal{L}^\Box$ and $\mathscr{M} = \langle W, R, V \rangle$, we say that $\mathscr{M}$ *satisfies* α if there is at least one world $w \in W$ such that $\mathscr{M}, w \Vdash \alpha$. We say that $\mathscr{M}$ is a *model of* α (alias α is *true* in $\mathscr{M}$), denoted $\mathscr{M} \Vdash \alpha$, if and only if $\mathscr{M}, w \Vdash \alpha$ for every world $w \in W$. Given a class (i.e., a collection) of models $\mathcal{M}$, we say that α is *valid* in $\mathcal{M}$, denoted $\models_{\mathcal{M}} \alpha$, if and only if every Kripke model $\mathscr{M} \in \mathcal{M}$ is a model of α. Given $\mathcal{K} \subseteq \mathcal{L}^\Box$ and $\alpha \in \mathcal{L}^\Box$, we say that $\mathcal{K}$ *locally entails* α in the class of models $\mathcal{M}$, denoted $\mathcal{K} \models_{\mathcal{M}} \alpha$, if and only if for every Kripke model $\mathscr{M} \in \mathcal{M}$ and every w in $\mathscr{M}$, if $\mathscr{M}, w \Vdash \beta$ for every $\beta \in \mathcal{K}$, then $\mathscr{M}, w \Vdash \alpha$. (When the class of models we are working with is clear from the context, we shall dispense with subscripts and just write $\models \alpha$ and $\mathcal{K} \models \alpha$.)

Here we shall assume the system of normal modal logic K, of which all the other normal modal logics are extensions. Semantically, K is characterised by the class of all Kripke models. Syntactically, K corresponds to the smallest set of sentences containing all propositional tautologies, all instances of the axiom schema $\mathsf{K} : \Box_i(\alpha \to \beta) \to (\Box_i\alpha \to \Box_i\beta)$, $1 \leq i \leq n$, and closed under the *rule of necessitation* below ($1 \leq i \leq n$):

$$(\text{RN})\ \frac{\alpha}{\Box_i\alpha}$$

For more details on modal logic, we refer the reader to the handbook by Blackburn et al. (Blackburn et al., 2006).

3. Towards a Logic for Expressing Defeasible Modalities

Recalling our discussion in the Introduction, we want to be able to state that a given sentence holds in *all* the relatively normal alternative worlds. This leads us to the definition of a 'weaker' version of the $\Box$ modality, which can be read as *defeasible necessity.* Through it we shall then be able to single out those normal situations that one cannot grasp via the classical $\Box$ modalities (recall the examples in the Introduction). Similarly, we want to be able to state that a given sentence holds in *at least one* relatively normal alternative world. This leads us to the definition of a stronger version of $\Diamond$, which, as we shall see, may be read as *distinct possibility.*

Example 1 below introduces the application scenario we shall use in the rest of the paper, with the purpose of illustrating more concretely the definitions and results that will follow.

Example 1. *We want to reason about a particular message, which may or may not be cyphered. If the message is cyphered, then it is safe. A given agent can encrypt, decrypt or transmit the message (i.e., broadcast it, making it public). Usually, if the message is public, then it is safe (because it presumably has been cyphered prior to transmission, which made it public). Transmitting the message may fail in producing its expected effects. Moreover, if the message is cyphered, then the agent (defeasibly) knows it is safe (since, plausibly, no one else can decrypt the message).*

We define a more expressive language than $\mathcal{L}^{\Box}$ by extending our modal language with a family of defeasible modal operators $\widetilde{\Box}_i$ and $\widetilde{\Diamond}_i$, $1 \leq i \leq n$ (called, respectively, 'flag' and 'flame'), where n is the number of classical modalities in the language. The sentences of the extended language are then recursively defined by:

$$\alpha \ ::= p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \Box_i\alpha \mid \widetilde{\Box}_i\alpha \mid \widetilde{\Diamond}_i\alpha$$

(As before, the other connectives are defined in terms of $\neg$ and $\wedge$ in the usual way, and $\top$ and $\bot$ are seen as abbreviations. It turns out that each $\widetilde{\Diamond}_i$ too is meant to be the dual of $\widetilde{\Box}_i$, as we shall see below.) With $\mathcal{L}^{\widetilde{\Box}}$ we denote the set of all sentences of such a richer language. Example 2 below provides some examples of $\mathcal{L}^{\widetilde{\Box}}$-sentences formalising the scenario introduced in Example 1.

Example 2. *Let $\mathcal{P} := \{\mathsf{p}, \mathsf{c}, \mathsf{s}\}$ be a set of propositions representing, respectively, "the message is public", "the message is cyphered" and "the message is safe". Let $\mathcal{A} := \{\mathsf{e}, \mathsf{d}, \mathsf{t}\}$ be a set of action names for the actions of "encrypting the message", "decrypting the message" and "transmitting the message". The following are examples of $\mathcal{L}^{\widetilde{\Box}}$-sentences: $\mathsf{c} \rightarrow \mathsf{s}$ ("if the message is cyphered, then it is safe"); $\Box_{\mathsf{e}}\mathsf{c} \wedge \Box_{\mathsf{d}}\neg\mathsf{c}$ ("encrypting cyphers the message and decrypting decyphers it"); $\widetilde{\Box}_{\mathsf{t}}\mathsf{p}$ ("a normal transmission of the message ensures that it is public") and $\mathsf{c} \rightarrow \widetilde{\Box}_{\mathsf{t}}\mathsf{s}$ ("if the message is cyphered, then a normal transmission ensures that the message remains safe").*

The question now is what an appropriate semantics for this language should be. Before providing a concrete answer, we shall briefly assess some natural candidates from the literature. (More details are provided in Section 7 on related work.)

Given the apparent similarities between the underlying intuition of $\widetilde{\Box}$ and $\widetilde{\Diamond}$ on the one hand and the semantic characterisation of constructs from conditional logics (Boutilier, 1994; Delgrande, 1988; Stalnaker, 1968), counterfactuals (Lewis, 1973) and some deontic logics (Hansson, 1969; Lewis, 1974) on the other, the obvious starting point would be to consider the semantic definitions from these frameworks. Indeed, all of them handle, in one way or another, operators that refer to "most typical" or "least abnormal" (or, in the case of deontic logic, "most ideal") situations.

Stalnaker's semantics for conditional logics (Stalnaker, 1968) is based on a *selection function* f which picks out the closest (most plausible) world to w satisfying a given sentence:

$$f : \mathcal{L}^{\Box} \times W \longrightarrow W$$

The obvious drawback of adopting such a definition in our context is that it assumes *uniqueness* of $f(\alpha, w)$, whereas we need the ability to single out possibly more than one most normal alternative world to the current one. For instance, a non-deterministic action may have more than one normal or expected outcome. To witness, when tossing a coin, there are two normal outcomes, viz. heads or tails, and an abnormal one, namely the coin standing on its edge.

In Lewis's systems of conditional and deontic logics (Lewis, 1973, 1974), the above mentioned uniqueness assumption is dropped and $f(\alpha, w)$ is defined as a subset of W instead. In principle, this should fit the bill. On the other hand, Lewis's constructions allow for a version of modus ponens for conditional statements:

$$\frac{\alpha,\ \alpha \Rightarrow \beta}{\beta}$$

Even though a case can be made for having such a principle in a purely conditional context, it becomes unwanted when interpreting conditionals as defeasible argument forms (Boutilier, 1994, p. 92), as can be seen in our scenario example:

Example 3. *Let the statement* $\mathsf{p} \Rightarrow \mathsf{s}$ *denote "normally, if the message is public, then it is safe". Now, if the message is meant to always be public, one would (wrongly) infer that it is always safe: from* p *and* $\mathsf{p} \Rightarrow \mathsf{s}$*, conclude* s*.*

Since our goal here is to move beyond defeasible conditionals but still making room for them (as we shall see in Section 5), we must bear in mind the side effects that adopting Lewis's approach would bring about. (We shall come back to general conditional logics *à la* Lewis in Section 7.)

Delgrande's (1987; 1988) approach also adopts the semantics of standard conditional logics and is based on a (general) selection function picking out the most normal worlds relative to the current one. In his setting, a conditional $\alpha \Rightarrow \beta$ holds at a world w if and only if the set of most normal α-worlds (relative to w) are also β-worlds. Besides the issues pointed out by Kraus et al. (1990), a problem with Delgrande's selection function f is that it is arbitrary in the sense that the selected worlds in $f(\alpha, w)$ need not satisfy α, which has the undesirable consequence that some α's may not normally imply themselves (Boutilier, 1994) — the so-called reflexivity property in KLM terms that we shall recall in Section 5. We shall come back to Delgrande's approach later on in the section on related work (Section 7).

Baltag and Smets's (2006; 2008) plausibility models capture some aspects of semantics that we have in mind, but their focus is on epistemic and doxastic reasoning, rather than on establishing a general framework, apt for reasoning e.g. about obligations or with ontologies in description logics (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2007), and integrating defeasible argument forms in the sense given by Kraus et al. (1990). (We shall return to the latter point in Section 5.)

Van Benthem (2010) outlines a modal logic of 'betterness', applicable to decision theory or game theory. Here, like in Baltag and Smets' approach, the preferences of each agent are explicit in the language in the form of a modal operator (the syntactic counterpart of the plausibility relation $\leq$). Obligations can also be expressed using the preference modality, but the resulting semantics of "α ought to be true" is then "α is true in all the best worlds". There is therefore no notion of defeasibility present as in our proposed reading of "α is true in all the *best alternative* worlds".

Of course, this is not to say that none of the aforementioned approaches are appropriate for our purposes here. At the end of the day, it remains a matter of finding a good compromise among aspects that are crucial from a knowledge representation and reasoning perspective. These include expressivity, intuitiveness, robustness, decidability, scalability and amenability to implementation, to name but a few. Here we shall strive for a semantic framework that (*i*) is expressive while still elegant and decidable, (*ii*) transfers smoothly to different contexts (and possibly to different logics), and (*iii*) also accounts for defeasible arguments of the form $\alpha \mathrel{|\!\sim} \beta$, i.e., that can easily be integrated with well-established approaches to non-monotonic reasoning such as the KLM one.

In this respect, we shall anchor our semantic constructions in the so-called preferential approach (Kraus et al., 1990; Shoham, 1988). The reason is threefold. First, it is acknowledged as one of the most comprehensive and successful frameworks for non-monotonic reasoning in the propositional case. Second, it transfers smoothly to more expressive languages such as modal logic (Britz et al., 2011a, 2012; Britz & Varzinczak, n.d., 2016b), description logics (Britz, Casini, Meyer, Moodley, & Varzinczak, 2013; Britz et al., 2008; Britz & Varzinczak, 2016a, 2017a, 2017b; Giordano, Gliozzi, Olivetti, & Pozzato, 2007, 2008; Giordano et al., 2009b; Giordano, Gliozzi, Olivetti, & Pozzato, 2012; Giordano et al., 2013, 2015), and others (Booth et al., 2015, 2012, 2013). Finally, it satisfies two fundamental desiderata in logic-based knowledge representation and reasoning, namely simplicity of the representation formalism and amenability to practical implementation (Casini, Meyer, Moodley, Sattler, & Varzinczak, 2015).

In the following, we provide the formal definition of our preferential semantics for modal logic and investigate some of the properties of the resulting framework for defeasible modalities.

4. A Preferential Kripke Semantics for Defeasible Modalities

In this section, we modify the constructions for preferential reasoning in modal logic as studied by Britz et al. (2011a; 2012) in the purely conditional case. We do so by enriching standard Kripke models with preference relations, instead of placing an ordering on states labeled by *pointed* Kripke models. Our starting point is therefore similar to the CT4O models of Boutilier (1994) and the plausibility models of Baltag and Smets (2006). (The differences will arise from the properties of our constructions and we shall point them out in more detail in Section 7.)

Definition 3 (Preferential Kripke Model). *A preferential Kripke model is a tuple $\mathscr{P} := \langle W, R, V, \prec\rangle$ where W is a (non-empty) set of possible worlds, $R := \langle R_1, \ldots, R_n\rangle$, where each $R_i \subseteq W \times W$ is an accessibility relation on W, $1 \leq i \leq n$, $V \colon W \longrightarrow \{0,1\}^{\mathcal{P}}$ is a valuation function, and $\prec \subseteq W \times W$ is a strict partial order (irreflexive and transitive) on W, satisfying the smoothness condition, i.e., $\prec$ has no infinitely descending chains.*

Given a preferential Kripke model $\mathscr{P} = \langle W, R, V, \prec\rangle$, we refer to $\mathscr{M} := \langle W, R, V\rangle$ as its associated standard Kripke model.

Definition 4 (Minimality w.r.t. $\prec$). *Let $\mathscr{P} = \langle W, R, V, \prec\rangle$ and let $W' \subseteq W$. Then $\min_{\prec} W' := \{w \in W' \mid$ there is no $w' \in W'$ such that $w' \prec w\}$, i.e., $\min_{\prec} W'$ denotes the minimal elements of W' with respect to $\prec$.*

The intuition behind the preference relation $\prec$ in a preferential Kripke model $\mathscr{P}$ is that worlds lower down in the order are *more preferred* (or deemed as being *more normal* (Booth et al., 2012; Boutilier, 1994)) than those higher up. Note that smoothness ensures $\min_{\prec} W' \neq \emptyset$, for every non-empty subset W' of the set of possible worlds in a preferential Kripke model.

It is worth pointing out that the preference relation in a preferential Kripke model, although a binary relation on W, is *not* to be seen as an accessibility relation. Indeed, the $\prec$-component in a preferential Kripke model has no counterpart in the syntax as each accessibility relation has (in the same way that possible wolds are not part of the syntax). As will be made clear later on, this need not be the case, but there are very good reasons for doing so.

We assume (for now) a single preference order across worlds in each preferential Kripke model, but of course Definition 3 can easily be generalised to a multi-preferential case. (This is particularly useful if one wants the ability to allow for several subjective orderings, like in a multi-agent context (Baltag & Smets, 2006).)

As an example, Figure 2 below depicts the preferential Kripke model $\mathscr{P}_1 := \langle W_1, R_1, V_1, \prec_1\rangle$, where $\langle W_1, R_1, V_1\rangle$ is as in Figure 1, and $\prec_1 := \{(w_1, w_2), (w_2, w_3), (w_1, w_3), (w_4, w_3)\}$, represented by the dashed arrows in the picture. (Note the direction of the dashed arrows, which point from more preferred to less preferred worlds.)

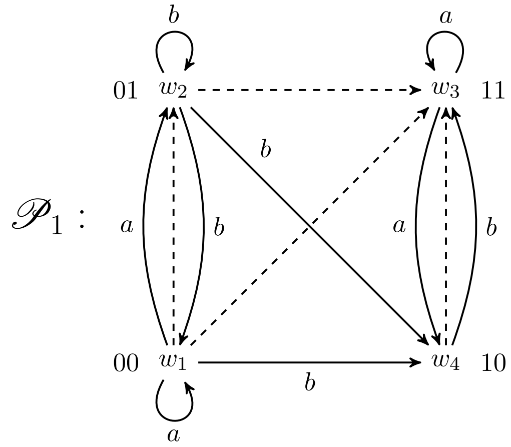


Figure 2. A Preferential Kripke model for $\mathcal{P} = \{p, q\}$ and two modalities.

If $\mathscr{P} = \langle W, R, V, \prec\rangle$ is a preferential Kripke model, $w \in W$ and $\alpha \in \mathcal{L}^{\Box}$ (i.e., α is a classical modal sentence), then $\mathscr{P}, w \Vdash \alpha$ if and only if $\mathscr{M}, w \Vdash \alpha$. With $[\![\alpha]\!]^{\mathscr{P}} := \{w \in W \mid \mathscr{M}, w \Vdash \alpha$, where $\mathscr{P} = \langle W, R, V, \prec\rangle\}$ we denote the set of possible worlds satisfying α (α-*worlds* for short) in $\mathscr{P}$. We say that α is *satisfiable* in $\mathscr{P}$ if $[\![\alpha]\!]^{\mathscr{P}} \neq \emptyset$, otherwise α is *unsatisfiable* in $\mathscr{P}$. We say that α is *true* in $\mathscr{P}$ (denoted $\mathscr{P} \Vdash \alpha$) if $[\![\alpha]\!]^{\mathscr{P}} = W$.

It is easy to see that the addition of the $\prec$-component preserves the truth of all (classical) modal sentences that are true in the underlying Kripke structure. That is, for every $\alpha \in \mathcal{L}^{\Box}$ and every $\mathscr{P} = \langle W, R, V, \prec\rangle$, $\mathscr{P} \Vdash \alpha$ if and only if $\langle W, R, V\rangle \Vdash \alpha$.

We can define *classes* of preferential Kripke models in the same way we do in the classical modal case. For instance, we can talk about the class of reflexive preferential Kripke models, in which the R-components are reflexive. We say that α is *valid* in the class $\mathcal{M}$ of preferential Kripke models if and only if α is true in every $\mathscr{P} \in \mathcal{M}$. Therefore, an immediate consequence of the observation in the preceding paragraph is that a classical modal sentence α is valid in the class $\mathcal{M}$ of preferential Kripke models if and only if it is valid in the corresponding class of Kripke models.

Armed with the notion of preferential Kripke models, we can provide a simple and intuitive semantics for our idea of defeasible modalities.

Definition 5 (Satisfaction Extended). *Let $\mathscr{P} = \langle W, R, V, \prec\rangle$ be a preferential Kripke model and $w \in W$.*

- *$\mathscr{P}, w \Vdash \widetilde{\Box}_i\alpha$ if and only if $\mathscr{P}, w' \Vdash \alpha$ for all w' such that $w' \in \min_{\prec} R_i(w)$;*
- *$\mathscr{P}, w \Vdash \widetilde{\Diamond}_i\alpha$ if and only if $\mathscr{P}, w' \Vdash \alpha$ for some w' such that $w' \in \min_{\prec} R_i(w)$.*

Given $\alpha \in \mathcal{L}^{\approx}$ (introduced on page 5) and preferential Kripke model $\mathscr{P} = \langle W, R, V, \prec\rangle$, as before, with $[\![\alpha]\!]^{\mathscr{P}}$ we denote the set of elements of W satisfying α. The notions of satisfaction in a preferential Kripke model, truth (in a model) and validity (in a class of preferential Kripke models) are extended to sentences of $\mathcal{L}^{\approx}$ in the obvious way.

The intuition behind a sentence like $\Box\!\!\!\!\sim_i\alpha$ is that α holds in the most normal of R_i-accessible worlds. $\Diamond\!\!\!\!\sim_i\alpha$ intuitively says that α holds in at least one such relatively normal accessible world. Example 4 below, which is a continuation of Example 2, illustrates more concretely these notions.

Example 4. *Consider the preferential Kripke model* $\mathscr{P}_2 := \langle W_2, R_2, V_2, \prec_2\rangle$*, which is depicted in Figure 3 below and where* $W_2 := \{w_i \mid 1 \leq i \leq 6\}$*,* $R_2 := \langle R_{\mathsf{e}}, R_{\mathsf{d}}, R_{\mathsf{t}}\rangle$*, with* $R_{\mathsf{e}} := \{(w_3, w_1), (w_4, w_2)\}$*,* $R_{\mathsf{d}} := \{(w_1, w_3), (w_2, w_4)\}$*,* $R_{\mathsf{t}} := \{(w_1, w_2),\ (w_2, w_2),\ (w_3, w_4),\ (w_3, w_5),\ (w_5, w_4),\ (w_5, w_6)\}$*,* V_2 *is the obvious valuation function, and* $\prec_2 := \{(w_1, w_3),\ (w_3, w_5),\ (w_1, w_5),\ (w_3, w_4),\ (w_1, w_4),\ (w_4, w_6),\ (w_3, w_6),\ (w_1, w_6),\ (w_2, w_4),\ (w_4, w_5),\ (w_2, w_5),\ (w_2, w_6)\}$*. (For the sake of readability, the transitive arrows of* $\prec_2$ *have been omitted from the picture.)*

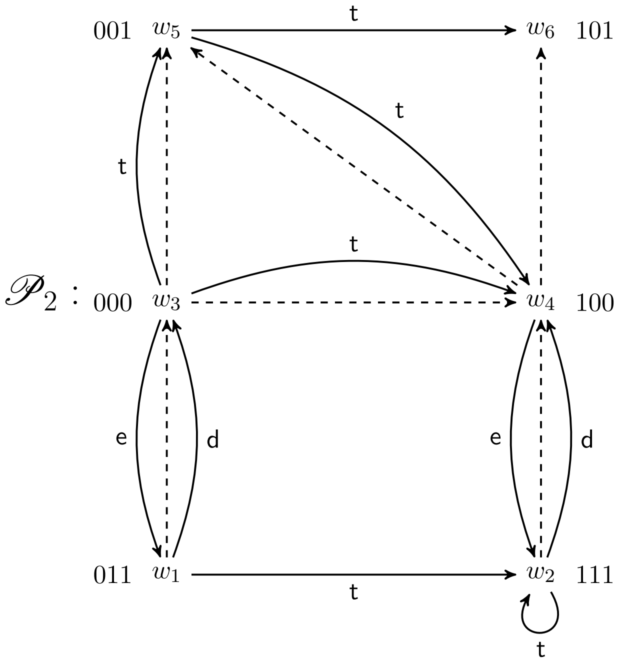


Figure 3. A Preferential Kripke model for $\mathcal{P} = \{\mathsf{p}, \mathsf{c}, \mathsf{s}\}$ and the action modalities $\mathcal{A} = \{\mathsf{e}, \mathsf{d}, \mathsf{t}\}$.

Given the preferential Kripke model $\mathscr{P}_2$ *in Figure 3, we can check that:*

- $\mathscr{P}_2 \Vdash \mathsf{c} \to \mathsf{s}$*: "if the message is cyphered, then it is safe";*
- $\mathscr{P}_2 \nVdash \Box\!\!\!\!\sim_{\mathsf{t}}\mathsf{s}$*: "it is not the case that (every) normal transmission of the message ensures that it is safe", since* $w_5 \notin [\![\Box\!\!\!\!\sim_{\mathsf{t}}\mathsf{s}]\!]^{\mathscr{P}_2}$*;*
- $\mathscr{P}_2 \Vdash \Box\!\!\!\!\sim_{\mathsf{t}}\mathsf{p}$*: "(any) normal transmission of the message ensures that it is public", but note that* $\mathscr{P}_2 \nVdash \Box_{\mathsf{t}}\mathsf{p}$*, since* $w_3 \notin [\![\Box_{\mathsf{t}}\mathsf{p}]\!]^{\mathscr{P}_2}$*.*

As mentioned before, in our enriched language, the preference relation is not explicit in the syntax. The meaning of the new modalities is informed by the preference relation, which nevertheless remains tacit outside the realm of defeasible modalities.[1] This stands in contrast to the approaches of Baltag and Smets (2006; 2008), Boutilier (1994), Britz et al. (2009) and Giordano et al. (2009a), which cast the preference relation as an extra modality in the object language. From a knowledge representation perspective, our approach has the advantage of hiding away some complex aspects of the semantics from the user (e.g. a knowledge engineer who will write down sentences in an agent's

[1] A similar approach is followed by Booth et al. (2015; 2012; 2013) in their extension of propositional logic to deal with 'typical' α-situations.

knowledge base and who will also have to maintain it during its life time). This speaks to the desiderata we mentioned at the end of Section 3.

We now turn our attention to some of the properties of preference-based defeasible modalities. We start by observing that, just as in the classical (i.e., non-defeasible) case, the defeasible modal operators $\widetilde{\Box}$ and $\widetilde{\Diamond}$ are the dual of each other, i.e., for $1 \leq i \leq n$:

$$\models \widetilde{\Box}_i \alpha \leftrightarrow \neg \widetilde{\Diamond}_i \neg \alpha \tag{1}$$

The following validities are also easy to verify ($1 \leq i \leq n$):

$$\models \widetilde{\Box}_i \bot \leftrightarrow \Box_i \bot, \quad \models \Diamond_i \top \leftrightarrow \widetilde{\Diamond}_i \top, \quad \models \widetilde{\Box}_i \top \leftrightarrow \top, \quad \models \widetilde{\Diamond}_i \bot \leftrightarrow \bot$$

The following is the $\widetilde{\Box}$-version of axiom schema K.

$$\models \widetilde{\Box}_i(\alpha \to \beta) \to (\widetilde{\Box}_i \alpha \to \widetilde{\Box}_i \beta) \tag{$\widetilde{\mathrm{K}}$}$$

The validity below is easy to verify:

$$\models \widetilde{\Box}_i(\alpha \land \beta) \leftrightarrow (\widetilde{\Box}_i \alpha \land \widetilde{\Box}_i \beta) \tag{$\widetilde{\mathrm{R}}$}$$

We also have $\models (\widetilde{\Box}_i \alpha \lor \widetilde{\Box}_i \beta) \to \widetilde{\Box}_i(\alpha \lor \beta)$, but not the converse, as can easily be checked.

The following validity testifies to the adequacy of our preferential semantics as an approach to defeasible modalities:

$$\models \Box_i \alpha \to \widetilde{\Box}_i \alpha \tag{$\widetilde{\mathrm{N}}$}$$

Intuitively, given $i = 1, \ldots, n$, where n is the number of modalities in the language, we want $\Box_i$ and $\widetilde{\Box}_i$ to be somehow 'tied together' in so far as one is the defeasible (respectively, the 'hard') version of the other. Schema ($\widetilde{\mathrm{N}}$) is in line with the commonly accepted principle that whatever is classically the case is also defeasibly so. (This is similar to what happens in KLM consequence relations, i.e., $\alpha \models \beta$ implies $\alpha \mid\!\sim \beta$ (Kraus et al., 1990), and in defeasible subsumption relations, i.e., $C \sqsubseteq D$ implies $C \underset{\sim}{\sqsubset} D$ (Britz et al., 2008).)

Given ($\widetilde{\mathrm{N}}$), it then follows that

$$\models \widetilde{\Diamond}_i \alpha \to \Diamond_i \alpha, \tag{2}$$

rendering support for our reading of $\widetilde{\Diamond}$ as *distinct* possibility.

It can easily be checked that in our preferential semantics, the standard rule of necessitation holds.[1] The following rule of *normal necessitation* (RNN) follows from RN together with Schema ($\widetilde{\mathrm{N}}$) above ($1 \leq i \leq n$):

$$(\mathrm{RNN})\ \frac{\alpha}{\widetilde{\Box}_i \alpha}$$

[1] Contrary to a version of preferential semantics for modal logic (Britz, Meyer, & Varzinczak, 2011b; Britz et al., 2012) based on states, which are labeled by pointed Kripke models, somewhat mimicking the original formulation by Kraus et al. (1990) in the propositional case.

From satisfaction of (1), ($\widetilde{\mathrm{K}}$) and ($\widetilde{\mathrm{R}}$), one can see that the logic of our defeasible modalities shares properties commonly characterising the so-called *normal* modal logics (Chellas, 1980). In particular, we have that the following rule holds:

$$\text{(NRK)}\ \frac{(\alpha_1 \land \ldots \land \alpha_n) \rightarrow \beta}{(\mathrel{\Box\!\!\!\sim}_i \alpha_1 \land \ldots \land \mathrel{\Box\!\!\!\sim}_i \alpha_n) \rightarrow \mathrel{\Box\!\!\!\sim}_i \beta}\ (n \geq 0)$$

The observant reader would have noticed that we assume there are as many defeasible modalities as there are classical ones. That is, for each $\Box_i$, a corresponding $\mathrel{\Box\!\!\!\sim}_i$ (its defeasible version) is assumed. Moreover, they are both linked together via Schema ($\widetilde{\mathrm{N}}$). In principle, from a technical point of view, nothing precludes us from having defeasible modalities with no corresponding classical version or the other way round. The latter is easily dealt with by simply not having $\mathrel{\Box\!\!\!\sim}_i$ for some i for which $\Box_i$ is present in the language. The former case, on the other hand, would require an elaboration of our semantics since the definition of satisfiability of $\mathrel{\Box\!\!\!\sim}$-sentences calls upon the accessibility relation R_i, associated with the $\Box_i$-modality. Even though one can make a case for only wanting the defeasible version of a given modality to be available in the syntax, it somewhat deviates from our stated aim of having defeasible *versions* of the (already existing) modalities in our language and we shall not investigate this further here.

The dependency between each (classical) modality and its defeasible counterpart is defined by a (fixed) preference order on worlds in the model. Since, by virtue of a motivated design choice (Section 3), the preference relation $\prec$ is not in the object language, clearly there can be no Hilbert-style axiomatisation of this dependency. One can get such an axiomatisation at the expense of adding new constructs to the language. For instance, by casting the preference order as a modality, one can axiomatise the relationship between $\mathrel{\Box\!\!\!\sim}_i$, $\mathrel{\Diamond\!\!\!\sim}_i$ and the preference order $\prec$, for each i. To this end, we may use, for example, the modal axiomatisation of the preference order of Britz et al. (2009), or one of Boutilier's (1994) modal systems. (We shall postpone the technical details until Section 7.) An axiomatisation then becomes possible at the expense of moving to a more expressive language (see the remark in the first paragraph after Example 4 and also the discussion in Section 7). Nevertheless, from a computational logic point of view, we shall suffice with the definition of a tableau-based decision procedure, which will be presented in Section 6.

5. Integrating Defeasible Modalities and KLM Conditionals

With defeasible modalities only we cannot directly express KLM-style conditionals of which the intuition is to capture a notion of defeasible argument form. Therefore, an obvious next step to the work presented here is the extension of $\mathcal{L}^{\mathrel{\Box\!\!\!\sim}}$ with a version of defeasible conditional, resulting in a framework allowing for the expression of both defeasible modalities and defeasible argument forms.[1]

We now enrich $\mathcal{L}^{\mathrel{\Box\!\!\!\sim}}$ with a defeasible implication connective $\rightsquigarrow$. The sentences of the extended language are then recursively defined by ($1 \leq i \leq n$):

$$\alpha \ ::= p \mid \neg\alpha \mid \alpha \land \alpha \mid \Box_i\alpha \mid \mathrel{\Box\!\!\!\sim}_i\alpha \mid \alpha \rightsquigarrow \alpha$$

(Again, the classical connectives are defined in terms of $\neg$ and $\land$ in the usual way, $\top$ and $\bot$ are seen as abbreviations, and each $\mathrel{\Diamond\!\!\!\sim}_i$ is the dual of $\mathrel{\Box\!\!\!\sim}_i$ — this being the reason

[1] In Section 7, we shall see how defeasible modalities can be used to simulate defeasible conditionals indirectly.

why we now drop $\Diamond\!\!\!\!\sim$ from the grammar.) A sentence of the form $\alpha \leadsto \beta$ should be read as "normally, if α, then β". With $\mathcal{L}^{\Box\!\!\!\!\sim+\leadsto}$ we denote the set of all sentences of such a richer language. Example 5 below provides some examples of $\mathcal{L}^{\Box\!\!\!\!\sim+\leadsto}$-sentences.

Example 5. *Let $\mathcal{P}$ and $\mathcal{A}$ be as in Examples 2–4. The following are examples of $\mathcal{L}^{\Box\!\!\!\!\sim+\leadsto}$-sentences: $\mathsf{p} \leadsto \mathsf{s}$ ("normally, if the message is public, then it is safe"); $\mathsf{p} \wedge \neg\mathsf{c} \leadsto \neg\mathsf{s}$ ("normally, if the message is public but not cyphered, then it is not safe") and $\top \leadsto \Box_{\mathsf{t}}\mathsf{s}$ ("normally, transmitting the message ensures that it is safe").*

Not surprisingly, our preferential Kripke semantics provides a natural way for interpreting $\leadsto$-sentences:

Definition 6 (Satisfaction Extended Further)**.** *Let $\mathscr{P} = \langle W, R, V, \prec\rangle$ be a preferential Kripke model and $w \in W$. For every $\alpha, \beta \in \mathcal{L}^{\Box\!\!\!\!\sim+\leadsto}$:*

- *$\mathscr{P}, w \Vdash \alpha \leadsto \beta$ if and only if $w \notin \min_{\prec}[\![\alpha]\!]^{\mathscr{P}}$ or $w \in [\![\beta]\!]^{\mathscr{P}}$.*

As before, the notions of satisfaction in a preferential Kripke model, truth (in a model) and validity (in a class of preferential Kripke models) are extended to sentences of $\mathcal{L}^{\Box\!\!\!\!\sim+\leadsto}$ in the obvious way.

The intuition of a sentence of the form $\alpha \leadsto \beta$ is that in those most normal situations in which α holds, β also holds. This is captured precisely by the following immediate consequence of Definition 6: $\mathscr{P} \Vdash \alpha \leadsto \beta$ if and only if $\min_{\prec}[\![\alpha]\!]^{\mathscr{P}} \subseteq [\![\beta]\!]^{\mathscr{P}}$. As an example, in the preferential model $\mathscr{P}_1$ of Figure 2, we have $\mathscr{P}_1 \Vdash \neg p \leadsto \Box_b \neg q$ (but note that $\mathscr{P}_1 \nVdash \neg p \to \Box_b \neg q$). We also have $\mathscr{P}_1 \Vdash p \leadsto \Diamond_b(q \wedge \Box_a p)$ and $\mathscr{P}_1 \nVdash \Box_a \neg p \leadsto q$ (from the latter follows $\mathscr{P}_1 \nVdash \Box_a \neg p \to q$).

The observant reader would have noticed that the semantics of our $\leadsto$-sentences differs from that of KLM-style conditionals in that here we adopt a 'local' notion of satisfaction, i.e., world driven, instead of model driven. This is the result of a number of choices we make in the present work: (*i*) the adoption of local entailment, rather than global entailment (cf. paragraph following Definition 2), (*ii*) the definition of $\leadsto$ at the *object* level, rather than at the meta-level, as it is the case with KLM conditionals, and (*iii*) the assumption of supraclassicality w.r.t. the underlying classical conditional.

There are strong arguments for each of these choices. About (*i*), local entailment is the standard notion of entailment adopted and motivated in, e.g., standard texts on modal logic such as Blackburn et al. (2006). With respect to (*ii*), including $\leadsto$ at the object level allows for nesting of defeasible conditionals (as motivated by Boutilier (1994)), and building complex modal expressions using different defeasible constructs. About (*iii*), the assumption of supraclassicality of the defeasible conditional is a cornerstone of non-monotonic reasoning in the KLM tradition.

It is worth noting that if only a classical underlying modal language is assumed (e.g. as in the work of Britz et al. (2011a; 2012)), then defeasible conditionals of the above form would still have the same intuition as mentioned in the Introduction. To witness, the statement $\Diamond\alpha \leadsto \Box\beta$ just says that "all normal worlds with an α-successor have only β-successors". That is, any $\leadsto$-sentence still refers only to normality in the premise, or, in this case, of the 'actual' world. In our enriched language, it becomes possible to exploit the different nuances resulting from where the normality spotlight is, as the following example, which is a continuation of Example 4, shows in a context involving actions, knowledge and conditionals.

Example 6. *Assume that A designates an agent's name and consider the preferential Kripke model $\mathscr{P}_3 := \langle W_3, R_3, V_3, \prec_3\rangle$, where $W_3 = W_2$, $V_3 = V_2$ and $\prec_3 = \prec_2$ as in $\mathscr{P}_2$ (Figure 3), and $R_3 := \langle R_{\mathsf{e}}, R_{\mathsf{d}}, R_{\mathsf{t}}, R_{\mathsf{A}}\rangle$, with R_{e}, R_{d}, R_{t} as in $\mathscr{P}_2$ and $R_{\mathsf{A}} := \{(w_1, w_1)$, (w_1, w_2), (w_2, w_2), (w_2, w_1), (w_5, w_5), (w_5, w_6), (w_6, w_6), $(w_6, w_5)\}$. Figure 4 depicts a* *representation of $\mathscr{P}_3$. (Again, in the picture we omit the transitive arrows of $\prec_3$.)*

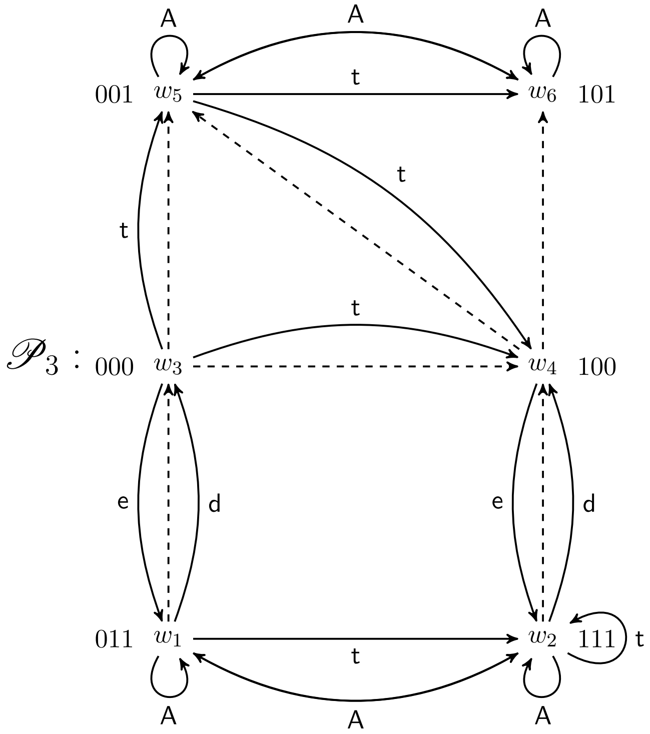


Figure 4. A Preferential Kripke model for $\mathcal{P} = \{\mathsf{p}, \mathsf{c}, \mathsf{s}\}$, the action modalities $\mathcal{A} := \{\mathsf{e}, \mathsf{d}, \mathsf{t}\}$ and the extra epistemic modality A. Note that transitivity of $\prec_3$ is implicit in the picture.

Given the preferential Kripke model $\mathscr{P}_3$ in Figure 4, one can check that:

- $\mathscr{P}_3 \Vdash \mathsf{p} \leadsto (\neg\mathsf{c} \leadsto \neg\mathsf{s})$*: "normally, if the message is public, then, normally, if it is not cyphered, it is not safe either";*
- $\mathscr{P}_3 \Vdash \Box_{\mathsf{A}}(\mathsf{p} \leadsto \mathsf{s})$*: "*$\mathsf{A}$ *knows that, normally, if the message is public, then it is safe";*
- $\mathscr{P}_3 \Vdash \mathsf{p} \leadsto \Box_{\mathsf{A}}\mathsf{s}$*: "normally, if the message is public, then* A *knows it is safe";*
- $\mathscr{P}_3 \Vdash \mathsf{p} \to \widetilde{\Box}_{\mathsf{A}}\mathsf{s}$*: "if the message is public, then* A *defeasibly knows that it is safe";*
- $\mathscr{P}_3 \Vdash \mathsf{p} \to \Box_{\mathsf{A}}(\top \leadsto \mathsf{s})$*: "if the message is public, then* A *knows that it is normally safe";*
- $\mathscr{P}_3 \Vdash \widetilde{\Box}_{\mathsf{t}}\Box_{\mathsf{A}}\mathsf{s}$*: "normal transmission of the message ensures that* A *knows it is safe". But note that* $\mathscr{P}_3 \not\Vdash \Box_{\mathsf{A}}\widetilde{\Box}_{\mathsf{t}}\mathsf{s}$*, since* $w_5 \notin [\![\Box_{\mathsf{A}}\widetilde{\Box}_{\mathsf{t}}\mathsf{s}]\!]^{\mathscr{P}_3}$.

The addition of a defeasible implication connective to $\mathcal{L}^{\widetilde{\Box}}$ represents a natural progression of work in the non-monotonic reasoning tradition, including the KLM framework (1990; 1992) and Boutilier's conditional logics of normality (1994). The following result shows that our defeasible implication connective $\leadsto$ behaves in a way that is commonly viewed as appropriate in the literature on non-monotonic reasoning. To be specific, $\leadsto$-sentences satisfy versions of the well-known basic KLM rationality postulates (Kraus et al., 1990), courtesy of our preferential Kripke semantics.

Proposition 1. *For every* $\alpha, \beta, \gamma \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ *and every preferential Kripke model* $\mathscr{P}$*:*

- $\mathscr{P} \not\Vdash \top \leadsto \bot$ *(Consistency)*
- $\mathscr{P} \Vdash \alpha \leadsto \alpha$ *(Reflexivity)*
- *If* $\models \alpha \leftrightarrow \beta$ *and* $\mathscr{P} \Vdash \alpha \leadsto \gamma$*, then* $\mathscr{P} \Vdash \beta \leadsto \gamma$ *(Left Logical Equivalence)*
- *If* $\mathscr{P} \Vdash \alpha \leadsto \beta$ *and* $\mathscr{P} \Vdash \alpha \leadsto \gamma$*, then* $\mathscr{P} \Vdash \alpha \leadsto \beta \wedge \gamma$ *(And)*
- *If* $\mathscr{P} \Vdash \alpha \leadsto \gamma$ *and* $\mathscr{P} \Vdash \beta \leadsto \gamma$*, then* $\mathscr{P} \Vdash \alpha \vee \beta \leadsto \gamma$ *(Or)*

- *If* $\mathscr{P} \Vdash \alpha \leadsto \beta$ *and* $\models \beta \rightarrow \gamma$, *then* $\mathscr{P} \Vdash \alpha \leadsto \gamma$ *(Right Weakening)*
- *If* $\mathscr{P} \Vdash \alpha \leadsto \beta$ *and* $\mathscr{P} \Vdash \alpha \leadsto \gamma$, *then* $\mathscr{P} \Vdash \alpha \wedge \gamma \leadsto \beta$ *(Cautious Monotony)*

Proof. See Appendix A.1. □

Furthermore, we have that both $\models \neg\alpha \leadsto (\alpha \leadsto \beta)$ and $\models \beta \leadsto (\alpha \leadsto \beta)$. This stands in contrast with the conditional logics of Stalnaker (1968) and Lewis (1973), where the exigence of avoiding the paradoxes of material implication was one of the main motivations for their introduction. However, note that the following global variants of the paradoxes of material implication do hold in the KLM framework:

$$\frac{p \mathrel{|\!\sim} \bot}{p \mathrel{|\!\sim} s} \qquad \frac{\neg s \mathrel{|\!\sim} \bot}{p \mathrel{|\!\sim} s} \tag{3}$$

We also have that for every $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ and every $\mathscr{P}$, $\mathscr{P} \Vdash \alpha$ if and only if $\mathscr{P} \Vdash \neg\alpha \leadsto \bot$. To see why, note that $\mathscr{P} \Vdash \alpha$ if and only if $[\![\alpha]\!]^{\mathscr{P}} = W$ if and only if $[\![\neg\alpha]\!]^{\mathscr{P}} = \emptyset$ if and only if $\min_{\prec}[\![\neg\alpha]\!]^{\mathscr{P}} = \emptyset$ if and only if $\min_{\prec}[\![\neg\alpha]\!]^{\mathscr{P}} \subseteq [\![\bot]\!]^{\mathscr{P}}$ if and only if $\mathscr{P} \Vdash \neg\alpha \leadsto \bot$.

We conclude this section by observing that $\leadsto$-sentences do not in general satisfy the Rational Monotony property below proposed by Lehmann and Magidor (1992).

- If $\mathscr{P} \Vdash \alpha \leadsto \beta$ and $\mathscr{P} \nVdash \alpha \leadsto \neg\gamma$, then $\mathscr{P} \Vdash \alpha \wedge \gamma \leadsto \beta$

We shall not develop this further here, but we point out that by restricting our semantics to *ranked Kripke models* (Britz et al., 2011a), i.e., preferential Kripke models in which the preference relation is a modular order,[1] we get a definition of $\leadsto$ satisfying all the rationality postulates. Since ranked Kripke models are a subclass of preferencial Kripke models, this would not affect the semantics of $\widetilde{\Box}$-sentences as studied in Section 4.

6. Tableau Calculus

In this section, we define a tableau calculus for reasoning with defeasible modalities and defeasible conditionals. The calculus is based on labeled sentences and on explicit accessibility relations (Goré, 1999).[2] As we shall see, it also makes use of an auxiliary structure of which the intention is to build a preference relation on possible worlds. (For a discussion on the differences between our tableau method and the one by Giordano et al. (2009a), which deals specifically with $\mathrel{|\!\sim}$-statements in a propositional setting, see the end of Section 7.)

Definition 7 (Labeled Sentence). *If* $n \in \mathbb{N}$ *and* $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$, *then* $n :: \alpha$ *is a labeled sentence.*

In a labeled sentence $n :: \alpha$, n is the *label*. (As we shall see, informally, the idea is that the label stands for some possible world in a Kripke model.)

Let $mod(\mathcal{L}^{\widetilde{\Box}+\leadsto})$ denote the set of all *classical modalities* of $\mathcal{L}^{\widetilde{\Box}+\leadsto}$. (Remember our assumption that we have as many defeasible modalities as we have classical ones and that, for a given i, both $\Box_i$ and $\widetilde{\Box}_i$ semantically depend on the same R_i. This explains why in the definition below it is enough to consider only the classical modalities in the construction of a structure which, intuitively, corresponds to the accessibility relations on worlds.)

[1] Given a set X, $\prec \subseteq X \times X$ is modular if and only if there is a ranking function $rk : X \longrightarrow \mathbb{N}$ such that for every $x, y \in X$, $x \prec y$ if and only if $rk(x) < rk(y)$. Note that modular orders can be obtained from total preorders by imposing anti-symmetry.

[2] Our exposition here follows that given by Varzinczak (2002) and Castilho et al. (1999; 2002).

Definition 8 (Skeleton)**.** *A skeleton is a function* $\Sigma : \mathit{mod}(\mathcal{L}^{\widetilde{\Box}+\rightsquigarrow}) \longrightarrow \mathscr{P}(\mathbb{N} \times \mathbb{N})$.

Informally, a skeleton maps modalities in the language to accessibility relations on the set of possible worlds.

Definition 9 (Preference)**.** *A preference relation* $\prec$ *is a binary relation on* $\mathbb{N}$.

As alluded to above, $\prec$ is meant to capture a preference relation on possible worlds. As we shall see below, like Σ, $\prec$ is built cumulatively through successive applications of the tableau rules we shall introduce.

Definition 10 (Branch)**.** *A branch is a tuple* $\langle \mathcal{S}, \Sigma, \prec \rangle$*, where* $\mathcal{S}$ *is a set of labeled sentences,* Σ *is a skeleton and* $\prec$ *is a preference relation.*

Definition 11 (Tableau Rule)**.** *A tableau rule is a rule of the form:*

$$\rho \ \ \frac{\mathcal{N} \ ; \ \Gamma}{\mathcal{D}_1 \ ; \ \Gamma'_1 \mid \ \ldots \ \mid \mathcal{D}_k \ ; \ \Gamma'_k}$$

where $\mathcal{N}; \Gamma$ *is the* numerator *and* $\mathcal{D}_1 \ ; \ \Gamma'_1 \mid \ \ldots \ \mid \mathcal{D}_k \ ; \ \Gamma'_k$ *is the* denominator.

Given a rule ρ, $\mathcal{N}$ represents one or more labeled sentences, called the *main sentences* of the rule, separated by ','. Γ stands for any additional *conditions* (on Σ or $\prec$) that must be satisfied for the rule to be applicable. In the denominator, each $\mathcal{D}_i$, $1 \leq i \leq k$, has one or more labeled sentences, whereas each Γ'_i is the additional conditions to be satisfied *after* the application of the rule (e.g. changes in the skeleton Σ or in the relation $\prec$). The symbol '|' indicates the occurrence of a *split* in the branch, i.e., a non-deterministic choice of the possible outcomes, each of which has to be explored.

Figure 5 presents the set of tableau rules for $\mathcal{L}^{\widetilde{\Box}+\rightsquigarrow}$. We say that a rule ρ is *applicable* to a branch $\langle \mathcal{S}, \Sigma, \prec \rangle$ if and only if $\mathcal{S}$ contains an instance of the main sentences of ρ and the conditions Γ of ρ are satisfied by Σ and $\prec$ (whenever appropriate). In the rules, we abbreviate $(n, n') \in \Sigma(i)$ as $n \xrightarrow{i} n'$, and $n' \in \Sigma(i)(n)$ as $n' \in \Sigma_i(n)$. With $n^\star, n'^\star, n''^\star, \ldots$ we denote labels that have not been used before (in the sequence of rule applications). For every α, we let $W^\alpha_\mathcal{S} := \{n \mid n :: \alpha \in \mathcal{S}\}$. Finally, with $n \in \min_\prec X$ we denote the fact that n is a minimal element in the set X, i.e., there is no $n' \in X$ such that $n' \prec n$.

The Boolean rules together with $(\Box_i)$ are as usual and need no explanation. Rule $(\widetilde{\Box}_i)$ propagates sentences in the scope of a defeasible necessity operator to the most preferred (with respect to $\prec$) of all accessible nodes. Rule $(\widetilde{\Diamond}_i)$ creates a preferred accessible node with the corresponding labeled sentences as content.

Rule $(\Diamond_i)$ replaces the standard rule for $\Diamond$-sentences and requires a more thorough explanation. Unlike in the defeasible version of the rule, namely $(\widetilde{\Diamond}_i)$, we cannot assume that there is a minimal accessible $\neg\alpha$-world. We therefore need to take into account the additional possibility that the accessible $\neg\alpha$-world is not minimal. (This has to be dealt with explicitly in order to ensure soundness of the algorithm.) Therefore, when creating a new accessible node, there are two possibilities: either (*i*) it is minimal (with respect to $\prec$) amongst all the accessible nodes, in which case the result is the same as that of applying Rule $(\widetilde{\Diamond}_i)$, or (*ii*) it is not minimal, in which case there must be a most preferred accessible node that is more preferred (with respect to $\prec$) than the newly created one. (This splitting is of the same nature as that in the $(\vee)$-rule, i.e., it fits the purpose of a proof by cases.)

Rules $(\rightsquigarrow)$ and $(\not\rightsquigarrow)$ take care of, respectively, $\rightsquigarrow$-sentences and their negations. Note that, in Rule $(\rightsquigarrow)$, the accessible preferred $n^\star$-world need not be minimal. It is not hard to see that this does not mean $\prec$ may fail the smoothness condition. Indeed, the finite depth of nesting of $\rightsquigarrow$ in the language ensures that no infinite chains of increasingly

$$(\bot)\ \frac{n :: \alpha,\ n :: \neg\alpha}{n :: \bot} \qquad (\neg)\ \frac{n :: \neg\neg\alpha}{n :: \alpha} \qquad (\wedge)\ \frac{n :: \alpha \wedge \beta}{n :: \alpha,\ n :: \beta} \qquad (\vee)\ \frac{n :: \neg(\alpha \wedge \beta)}{n :: \neg\alpha \mid n :: \neg\beta}$$

$$(\Box_i)\ \frac{n :: \Box_i\alpha\ ;\ n \xrightarrow{i} n'}{n' :: \alpha} \qquad (\widetilde{\Box}_i)\ \frac{n :: \widetilde{\Box}_i\alpha\ ;\ n \xrightarrow{i} n',\ n' \in \min_{\prec} \Sigma_i(n)}{n' :: \alpha}$$

$$(\widetilde{\Diamond}_i)\ \frac{n :: \neg\widetilde{\Box}_i\alpha}{n'^{\star} :: \neg\alpha\ ;\ n \xrightarrow{i} n'^{\star},\ n'^{\star} \in \min_{\prec} \Sigma_i(n)} \quad (\Diamond_i)\ \frac{n :: \neg\Box_i\alpha}{n'^{\star} :: \neg\alpha\ ;\ \Gamma'_1 \mid n'^{\star} :: \neg\alpha\ ;\ \Gamma'_2},\ \text{where:}$$

$$\Gamma'_1 = \{n \xrightarrow{i} n'^{\star},\ n'^{\star} \in \min_{\prec} \Sigma_i(n)\}\ \text{and}$$
$$\Gamma'_2 = \{n \xrightarrow{i} n'^{\star},\ n \xrightarrow{i} n''^{\star},\ n''^{\star} \prec n'^{\star},\ n''^{\star} \in \min_{\prec} \Sigma_i(n)\}$$

$$(\leadsto)\ \frac{n :: \alpha \leadsto \beta}{n :: \neg\alpha \mid n^{\star} :: \alpha\ ;\ n^{\star} \prec n \mid n :: \beta} \qquad (\not\leadsto)\ \frac{n :: \neg(\alpha \leadsto \beta)}{n :: \alpha,\ n :: \neg\beta\ ;\ n \in \min_{\prec} W^{\alpha}_{\mathcal{S}}}$$

$$(\bot_{\prec})\ \frac{n \in \min_{\prec} W^{\alpha}_{\mathcal{S}},\ n' \prec n \text{ for some } n' \in W^{\alpha}_{\mathcal{S}}}{n :: \bot}$$

Figure 5. Tableau rules for defeasible modalities and conditionals.

preferred worlds can arise. Also, since $n^{\star}$ is not accessible from any other world in the partially constructed model, explicit mention of minimality (as in the case of the $\Diamond_i$-rule discussed above) is not required.

Finally, Rule $(\bot_{\prec})$ performs a supplementary (meta-) consistency check based on the preference relation and of which the explanation speaks for itself.

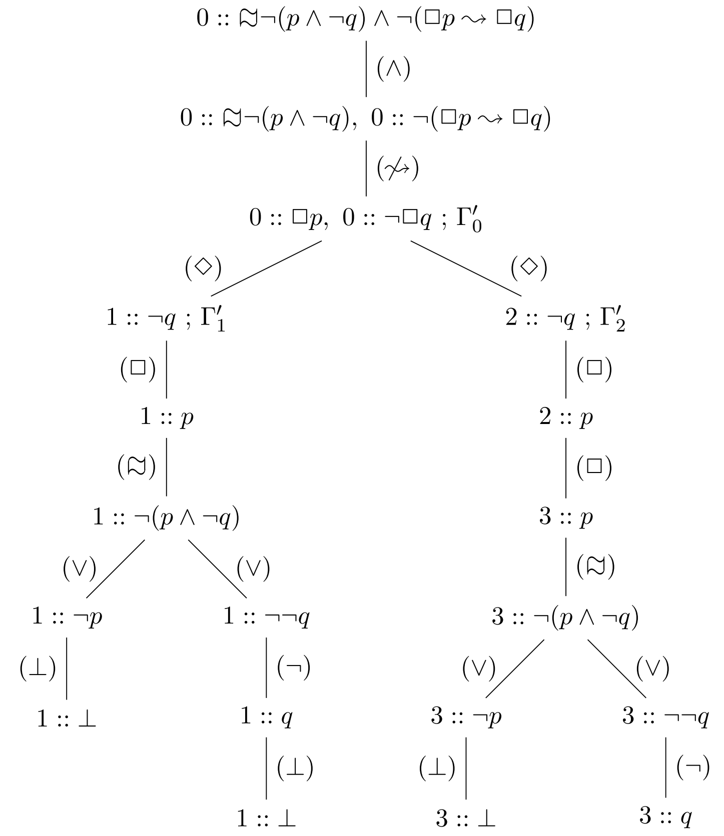
Definition 12 (Tableau). *A tableau $\mathcal{T}$ for $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ is the limit of a sequence $\mathcal{T}^0$, …, $\mathcal{T}^n$, … of sets of branches where the initial $\mathcal{T}^0 := \{\langle\{0 :: \alpha\}, \emptyset, \emptyset\rangle\}$ and every $\mathcal{T}^{i+1}$ is obtained from $\mathcal{T}^i$ by the application of one of the rules in Figure 5 to some branch $\langle\mathcal{S}, \Sigma, \prec\rangle \in \mathcal{T}^i$. Such a limit is denoted $\mathcal{T}^{\infty}$.*

Here we make the so-called *fairness assumption*: any rule that *can* be applied *will* eventually be applied, i.e., the order of rule applications is not relevant. Moreover, we assume that no rule is applied more than once to the same instance of a labeled sentence on the same branch. We say a tableau is *saturated* if no rule is applicable to any of its branches.

Definition 13 (Closed Tableau). *A branch $\langle\mathcal{S}, \Sigma, \prec\rangle$ is closed if and only if $n :: \bot \in \mathcal{S}$ for some n. A saturated tableau $\mathcal{T}$ for $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ is closed if and only if all its branches are closed. (If $\mathcal{T}$ is not closed, then we say that it is an* open tableau*.)*

For an example construction of a tableau, assume an underlying monomodal language and consider the sentence $\alpha := \widetilde{\Box}(p \to q) \to (\Box p \leadsto \Box q)$, which is not preferentially valid. Figure 6 below depicts the (open) tableau $\mathcal{T}_1$ for $\neg\alpha := \widetilde{\Box}\neg(p \wedge \neg q) \wedge \neg(\Box p \leadsto \Box q)$.

From the open tableau $\mathcal{T}_1$ shown in Figure 6, we extract the preferential Kripke model $\mathscr{P}_{\mathcal{T}_1}$ depicted in Figure 7. (In Figure 7, the understanding is that $3 \prec 2$ and that 0 is *incomparable* with respect to $\prec$ to the other possible worlds.)



$\Gamma'_0 =$ set 0 as $\min_{\prec} W_{\mathcal{S}}^{\Box p}$
$\Gamma'_1 =$ add $(0, 1)$ to Σ and set 1 as $\min_{\prec} \Sigma(0)$
$\Gamma'_2 =$ add $(0, 2)$ and $(0, 3)$ to Σ, $(3, 2)$ to $\prec$ and set 3 as $\min_{\prec} \Sigma(0)$

Figure 6. Visualisation of the open tableau $\mathcal{T}_1$ for $\approx\neg(p \wedge \neg q) \wedge \neg(\Box p \leadsto \Box q)$.

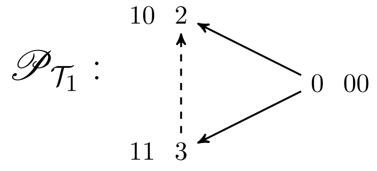


Figure 7. Preferential Kripke model $\mathscr{P}_{\mathcal{T}_1}$ constructed from Figure 6.

We are now ready to state the main result of this section.

Theorem 1. *The tableau calculus for $\mathcal{L}^{\approx+\leadsto}$ is sound and complete with respect to our modal preferential semantics.*

Proof. See Appendix A.2. □

Moreover, our tableau calculus provides us with a decision procedure for $\mathcal{L}^{\approx+\leadsto}$:

Theorem 2. *The tableau calculus for $\mathcal{L}^{\approx+\leadsto}$ terminates.*

Proof. It can easily be checked that in the construction of the tableau there is only a finite number of distinct states since every sentence generated by the application of a rule is a sub-sentence of the original one. □

We end this section by noting that the addition of both $\widetilde{\Box}$ and $\leadsto$ to the underlying modal language does not affect the space complexity of the resulting tableaux.

Theorem 3. *Satisfiability checking for* $\mathcal{L}^{\widetilde{\Box}+\leadsto}$ *is* PSPACE*-complete.*

Proof. It is well-known that satisfiability checking for classical modal logic K and K_n are both PSPACE-complete (Halpern & Moses, 1992; Ladner, 1977). If the sentence at the root of the tableau is α, and $|\alpha| = m$, i.e., m is the number of symbols occurring in α, then the space requirement for each label is at most $O(m)$. Since there is a saturated tableau with depth at most $O(m^2)$, the total space requirement is $O(m^3)$. □

In summary, in spite of the additional expressivity brought in by the introduction of preferential versions of modal operators and of a defeasible conditional, we remain in the same complexity class as that of the modal logic we started off with.

7. Related Work

In what follows, we shall restrict our discussion to a representative selection of influential related research efforts.

Conditional Logics

We start with a remark on a possible translation between our $\widetilde{\Box}$-logic and conditional logic *à la* Lewis. Assume a *multi-conditional* language, i.e., assume we have a collection of operators $\Rightarrow_i$, $1 \leq i \leq n$. For each w, let the selection function $f(\cdot)$ pick out the minimal of the accessible worlds from w. Then each $\widetilde{\Box}_i\alpha$ can be captured as the statement $\top \Rightarrow_i \alpha$. (This of course does not mean that the $\widetilde{\Box}$-logic is trivial since smoothness and minimality still have to be axiomatised.) For the other direction, i.e., for an embedding of conditionals in a $\widetilde{\Box}$-language, for each $\alpha \in \mathcal{L}^{\Box}$, let $R_\alpha := \{(w, w') \mid w \Vdash \alpha \text{ and } w' \Vdash \alpha\}$. Then the conditional $\alpha \Rightarrow \beta$ is captured by the sentence $\widetilde{\Box}_\alpha\beta$. Note that this mapping still does not define each of the (classical) $\Box_i$, which would have to be defined separately, and semantically linked to the corresponding $\widetilde{\Box}_i$.

The exigence of avoiding the paradoxes of material implication was one of the main motivations for the introduction of conditional logics by Stalnaker and Lewis. Our integration of defeasible implication $\leadsto$ with our language of defeasible modalities is rather a natural progression of work in the non-monotonic reasoning tradition. To the best of our knowledge, the first attempt to formalise a notion of relative normality in the context of defeasible reasoning was that of Delgrande (1987; 1988) in which a conditional logic of normality is defined. Given the relationship between the general constructions on which we base our work and those by Kraus et al. (1990), most of the remarks in the comparison made by Lehmann and Magidor (1992, Section 3.7) are applicable in comparing Delgrande's approach to ours and we do not repeat them here. We note though that, like Kraus et al. and Boutilier, Delgrande focuses on defeasibility of argument forms rather than modes of reasoning as we studied here. Contrary to them, and as we have seen in Section 3, Delgrande adopts the semantics of standard conditional logics (Chellas, 1980, Chapter 10). In his setting, a conditional $\alpha \Rightarrow \beta$ holds at a world w if and only if the set of most normal α-worlds (relative to w) are also β-worlds. We can capture Delgrande's conditionals in our approach with $\widetilde{\Box}$-sentences of the form $\widetilde{\Box}(\alpha \to \beta)$ in the class of $\mathsf{S5}$ preferential Kripke models.

Boutilier's (1994) expressive conditional logics of normality act as unifying framework for a number of conditional logics, including those of Delgrande and Kraus et al. The main difference between his approach and ours is in whether the underlying preference order alters the meaning of modalities or not. Boutilier's conditional is defined directly from a preference order in a bi-modal language, but the meanings of any additional, independently axiomatised, modalities are not influenced by the preference order. Our defeasible modalities correspond to a modification of the other (classical) modalities using the preference relation.

Makinson's Conditional Obligations

In his comparative perspective on minimality, Makinson (1993) outlined a modal reading of conditional obligations rooted in the work of Hansson (1969), Lewis (1974) and others. Although he does not develop the formal details of the proposal, there are clear connections with a deontic reading of our defeasible modalities, both technically and philosophically. We therefore first sketch his proposal, and then link it to our work.

In Makinson's proposal, the (indefeasible) conditional obligation $O(\beta/\alpha)$ is true in a world w if and only if, for every world w' future to w in which α is true, β is true in all of the best among the worlds w'' that are in turn accessible from w'. Leaving aside the conditional and temporal aspects represented by α for a moment, this renders the following simplified 'customary' reading of unconditional obligation: $O\beta$ is true in w if and only if β is true in the best of the worlds accessible from w. Now, this appears to coincide with $\Box\!\!\!\!\sim$ when it is interpreted as defeasible obligation, but one has to recall that the conventional 'best of the worlds' presumably referred to here are the 'ideal worlds' selected by a choice function as found in counterfactual conditionals. The best of the worlds here therefore have no bearing on the defeasibility of the obligation; rather, it is part of the conventional stance on obligations referred to by Makinson, which can be given a modal treatment. Of course, one can reinterpret the original meaning by placing a preference order on worlds, apart from the accessibility relation corresponding to the deontic modality. The best of the accessible worlds would then coincide with our notion of defeasible obligation. In this sense, our work represents a natural progression of, e.g. Lewis's 'best of a bad lot', but it is incorrect to say that it amounts to the same thing.

Makinson then suggests bringing in a defeasible aspect by placing a normality order on the temporal modality. So, instead of considering "every world w' future to w in which α is true", we have "the most normal among the worlds w' future to w in which α is true". This restriction yields, in our setting, a defeasible temporal modality. That is, given the customary temporal modality F, we may form the defeasible modality $\Box\!\!\!\!\sim_F$, with $\Box\!\!\!\!\sim_F\alpha$ true in w if and only if α is true in all the best of the worlds w' amongst those in the future of w.

The defeasibility suggested by Makinson is therefore not in the obligation itself, but rather in the temporal aspect of the conditional. However, technically speaking, the idea of a normality order interacting with a modality to introduce defeasibility is already present here. As Makinson (1993) points out: "defeasibility can be expressed by introducing the concept of normality as an additional ingredient of the satisfaction rule." Our proposal naturally incorporates this idea. That is, we may introduce both a temporal and a deontic modality and form both their defeasible counterparts. This would then yield a logic in which we can express both defeasible conditionality and defeasible obligations. These two aspects of defeasibility in deontic reasoning remains to be investigated in more detail, whether in the context of dyadic or monadic deontic logic.

Plausibility Models

Baltag and Smets (2006; 2008) also employ preference orders to refer to the normality (or plausibility) of accessible worlds. However, their aims and resulting semantics differ from ours in some key aspects beyond those mentioned in Section 3. On one hand, they define multi-agent epistemic and doxastic *plausibility models* that are similar to our preferential Kripke models, but each accessibility relation is induced by a corresponding preference order and linked to an agent whose beliefs are determined by what the agent deems epistemically possible (contrast this with our classical modalities, which are defined independently from the preference order). Minimality, or *doxastic appearance*, is therefore determined relative to an epistemic context, which is induced as an equivalence relation on worlds. This results in modalities of knowledge, (conditional) belief and safe belief that are complementary to epistemic and doxastic versions of our defeasible modalities.

To see how this is the case, let $\mathscr{M}_p := \langle W, \sim, \leq, V\rangle$ be a plausibility model, where W and V are as usual, $\sim$ is an equivalence relation on W and $\leq$ is a plausibility relation on worlds. The latter induces a *safe belief* operator $\Box_{\leq}$ of which the semantics is given by:

$$\mathscr{M}_p, w \Vdash \Box_{\leq}\alpha \text{ if and only if for every } w' \text{ such that } w' \leq w,\ \mathscr{M}, w' \Vdash \alpha$$

In other words, their $\Box_{\leq}\alpha$ is true in a world w if and only if α is true in *all better* accessible worlds from w. Contrast this with our $\Box\!\!\!\!\sim\alpha$, which is true in w if and only if α is true in *all best* accessible worlds. Moreover, in their case the ‘current’ w is always amongst the selected worlds, whereas in our case it need not be (even if $\mathsf{S5}$ is assumed, since there may be accessible worlds that are more preferred than the current one). An analysis of the philosophical implications of each of these two choices (“all the better” v. “all the best”) in specific contexts, as well as a combination thereof, is a possible avenue of further exploration.

Vague Modalities

The present paper defines modal constructs suitable for reasoning about defeasible modes of inference. More generally, one may consider the definition of suitable modal constructs for reasoning about vague notions such as ‘generally’, ‘rarely’ or ‘most’ (Askounis, Koutras, & Zikos, 2012). Such was the aim of Veloso et al. (2009) in their extension of Kripke semantics to incorporate generalised modal operators for vague notions on accessibility relations. Their framework is very broad, without the restrictions imposed by a preference order on possible worlds, but their aims are aligned with ours in that they present a direct modal treatment of vague assertions. Briefly, they extend standard modal logic with a new modal operator ∇, whose intended meaning is that $\nabla\alpha$ holds at a state w if and only if the set of states reachable from w where α holds is an ‘important set’. These sets deemed as important are defined by a *complex* K over a frame $\langle W, R\rangle$, which is a function mapping each $w \in W$ to a family of subsets $K(w) \subseteq \mathscr{P}(R(w))$. We then have

$$\mathscr{M}, w \Vdash \nabla\alpha \text{ if and only if } \{w' \mid (w, w') \in R \text{ and } \mathscr{M}, w' \Vdash \alpha\} \in K(w).$$

It follows that, in the semantics of the language $\mathcal{L}^{\Box\!\!\!\!\sim}$, if we define the modality ∇ by letting $K(w) := \{X \mid \min_{\prec} R(w) \subseteq X\}$, then $\mathscr{M}, w \Vdash \Box\!\!\!\!\sim\alpha$ if and only if $\mathscr{M}, w \Vdash \nabla\alpha$.

Not every complex K corresponds to some preference order on W, hence we do not have a converse result. Some vague notions that can be captured using ∇ are therefore not expressible using $\Box\!\!\!\!\sim$. Additional structure may be imposed on K, for example, by considering only families of ultra-filters as range. The structure provided by the preference order and principal filter generated by $\min_{\prec} R(w)$ and defined above thus define a useful class of defeasible modalities that fits into the broader framework for vague notions and modalities.

Preferential Approaches

Booth et al. (2015; 2012; 2013) introduce an operator with which one can refer directly in the object language to those *most typical* situations in which a given sentence is true. For instance, in their enriched language, a sentence of the form $\bullet\alpha$ refers to the 'most typical' α-worlds in a semantics similar to ours. One of the advantages of such an extension is the possibility to make statements of the kind "all normal α-worlds are normal β-worlds", thereby shifting the focus of normality from the antecedent by also allowing us to talk about normality in the consequent. This additional expressivity can also be obtained by the addition of the modality $\Box$ of Modular Gödel-Löb logic to express normality syntactically (Britz et al., 2009; Giordano et al., 2009a):

$$\bullet\alpha := \alpha \land (\Box_{\prec}\neg\alpha) \tag{4}$$

The modality $\Box_{\prec}$ in (4) above refers to the preference relation $\prec$ seen as an extra accessibility relation.

Despite the gain in expressivity, both the proposals by Booth et al. (2012) and Britz et al. (2009) remain propositional in nature in that the only modality allowed is the one with semantics determined by the preference order. Recently, Britz et al. (2011a; 2012) extended propositional preferential reasoning to the modal case, but the modalities under consideration there remain classical — their meaning remains as in propositional modal logic. This is so in spite of the underlying preferential semantics which is adopted to deal with conditional statements of the form $\alpha \mid\!\sim \beta$, where α and β are (classical) modal sentences.

The similarities between the tableau method we presented here and the one by Giordano et al. (2009a) are largely superficial. First, our preferential semantics counts as a proper generalisation of the KLM approach to full modal logic, whereas theirs is an embedding of propositional KLM consequence relations in a language enriched with a modality to represent the conditional $\mid\!\sim$. Second, again, in their approach the preference relation is explicit and cast as an additional modality, requiring a special tableau rule to deal with it. Here the preference relation is not present in the syntax and materialises only in the inner workings of our semantic tableau method.

8. Concluding Remarks

The main contribution of the present paper is the provision of a natural, simple and intuitive framework within which to represent defeasible modes of inference. The defeasible modalities we introduced here refer to the relative normality of *accessible worlds*, unlike characterisations of normality (Booth et al., 2012, 2013; Boutilier, 1994; Giordano et al., 2007, 2008, 2009a, 2009b), which refer to the relative normality of worlds in which a given sentence is true, or versions of $\mid\!\sim$ (Kraus et al., 1990; Lehmann & Magidor, 1992), which refer to the relative normality of the worlds in which the premise is true.

In our logic, new operators are introduced that enable the normality-based relationship among worlds to be implicitly indicated. From a knowledge representation perspective, this is an appealing feature. Indeed, reasoning with $\mathrel{\Box\!\!\!\sim}$-sentences is much easier than with other frameworks because the preference relation on worlds is implicit, in a similar way that reasoning with temporal logic is easier than with its translation into predicate calculus, as the relationships among time points are implicit.

Moreover, we have seen that in order for us to capture the semantics of $\widetilde{\Box}$-sentences in standard conditional logics we would require the addition of an explicit preference relation on worlds, all standard modalities we want to work with and a suitably defined conditional for each modality in the language. Our contention here is that this route would hardly simplify matters.

If instead we do want to internalise the preference relation in the object language, then by also enriching our classical modal language with converse modalities and nominals (Blackburn, de Rijke, & Venema, 2001), it turns out $\widetilde{\Box}$ can be given an entirely classical treatment as follows:

$$\widetilde{\Box}\alpha := \bigvee_{o\in\mathcal{O}} (o \wedge \Box(\neg\alpha \rightarrow \Diamond_{\prec}(\alpha \wedge \check{\Diamond} o))) \tag{5}$$

where $\Diamond_{\prec}$ is the dual of the modality characterising the preference relation (Britz et al., 2009), $\check{\Diamond}$ is the converse of $\Diamond$ and $\mathcal{O}$ is a set of nominals. Then $\widetilde{\Box}\alpha$ is true at a world w in a (hybrid) Kripke model if and only if w is the denotation of some nominal $o \in \mathcal{O}$ and every $\neg\alpha$-world that is accessible from w is less normal than some α-world which is accessible from w. (Of course, besides ensuring that each nominal is interpreted as at most one possible world one also has to make sure that each possible world in a Kripke model is the denotation of some nominal $o \in \mathcal{O}$. This is warranted in the class of *named* models (Blackburn et al., 2001, pp. 439–447).)

The definition in (5) has the inconvenience of requiring infinitary disjunctions (Karp, 1964) in the object language. We can replace (5) with the following axiom schema:

$$\text{(F)}\ @_o\widetilde{\Box}\alpha \leftrightarrow @_o\Box(\Box_{\prec}\neg\check{\Diamond} o \rightarrow \alpha)$$

As mentioned earlier, making use of such a machinery takes us to an unnecessarily more expressive language. Note though that complexity-wise we remain in the same class — satisfiability in the basic hybrid logic like the one briefly sketched above is PSPACE-complete (Blackburn et al., 2001, Theorem 7.21).

Here we have investigated the case where a single preference order on worlds is assumed. As we have seen, this fits the bill in capturing defeasibility of action effects or obligations, where an ‘objective’ or commonly agreed upon notion of normality can be justified. When moving to defeasible notions of knowledge or belief, though, a multi-preference based approach seems to be more appropriate, as agents may have different views on which worlds are more normal than others, i.e., preferences become *subjective* or at least relative to an agent (Baltag & Smets, 2008).

In this paper, we have investigated defeasible modalities in the system K. Our basic framework paves the way for exploring similar notions of defeasibility and additional properties in specific systems of modal logics.[1] Once this is in place, we will be able to investigate further applications of defeasible modalities in e.g. dynamic epistemic logic (Ditmarsch, Hoek, & Kooi, 2007) as well as in other similarly structured logics, such as description logics (Baader et al., 2007). We have recently investigated preferential versions of role restrictions in DLs along the lines of our preferential modalities (Britz, Casini, Meyer, & Varzinczak, 2013), showing that our definitions are also fruitful in the formalisation of different notions of defeasibility in ontologies.

[1] We have already caught a glimpse of this in Example 6, where an S5-modality for knowledge is assumed.

Finally, entailment in the modal logics obtained through the addition of $\mathrel{\Box\!\!\!\!\sim}$ is monotonic. In a broader defeasible reasoning context, though, a case can be made for nonmonotonic entailment relations such as those studied by Booth et al. (2015) in a propositional setting and by Giordano et al. (2013; 2015) in description logics.

Acknowledgements

This work is based on research supported in part by the National Research Foundation of South Africa (Grant Numbers 103345 and 85482).

References

Askounis, D., Koutras, C., & Zikos, Y. (2012). Knowledge means 'all', belief means 'most'. In L. Fariñas del Cerro, A. Herzig, & J. Mengin (Eds.), *Proceedings of the 13th european conference on logics in artificial intelligence (JELIA)* (p. 41-53). Springer.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (Eds.). (2007). *The description logic handbook: Theory, implementation and applications* (2nd ed.). Cambridge University Press.

Baltag, A., & Smets, S. (2006). Dynamic belief revision over multi-agent plausibility models. In W. van der Hoek & M. Wooldridge (Eds.), *Proceedings of LOFT* (p. 11-24). University of Liverpool.

Baltag, A., & Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, & M. Wooldridge (Eds.), *Logic and the foundations of game and decision theory (LOFT7)* (p. 13-60). Amsterdam University Press.

Benthem, J. (2010). *Modal logic for open minds.* Center for the Study of Language and Information.

Blackburn, P., Benthem, J., & Wolter, F. (2006). *Handbook of modal logic.* Elsevier North-Holland.

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic.* Cambridge University Press.

Booth, R., Casini, G., Meyer, T., & Varzinczak, I. (2015). On the entailment problem for a logic of typicality. In *Proceedings of the 24th international joint conference on artificial intelligence (IJCAI).*

Booth, R., Meyer, T., & Varzinczak, I. (2012). PTL: A propositional typicality logic. In L. Fariñas del Cerro, A. Herzig, & J. Mengin (Eds.), *Proceedings of the 13th european conference on logics in artificial intelligence (JELIA)* (p. 107-119). Springer.

Booth, R., Meyer, T., & Varzinczak, I. (2013). A propositional typicality logic for extending rational consequence. In E. Fermé, D. Gabbay, & G. Simari (Eds.), *Trends in belief revision and argumentation dynamics* (Vol. 48, p. 123-154). King's College Publications.

Boutilier, C. (1994). Conditional logics of normality: A modal approach. *Artificial Intelligence*, *68*(1), 87-154.

Britz, K., Casini, G., Meyer, T., Moodley, K., & Varzinczak, I. (2013). *Ordered interpretations and entailment for defeasible description logics* (Tech. Rep.). CAIR, CSIR Meraka and UKZN, South Africa. Retrieved from `http://tinyurl.com/cydd6yy`

Britz, K., Casini, G., Meyer, T., & Varzinczak, I. (2013). Preferential role restrictions. In *Proceedings of the 26th international workshop on description logics* (p. 93-106).

Britz, K., Heidema, J., & Meyer, T. (2008). Semantic preferential subsumption. In J. Lang & G. Brewka (Eds.), *Proceedings of the 11th international conference on principles of knowledge representation and reasoning (KR)* (p. 476-484). AAAI Press/MIT Press.

Britz, K., Heidema, J., & Meyer, T. (2009). Modelling object typicality in description logics. In A. Nicholson & X. Li (Eds.), *Proceedings of the 22nd australasian joint conference on artificial intelligence* (p. 506-516). Springer.

Britz, K., Meyer, T., & Varzinczak, I. (2011a). Preferential reasoning for modal logics. *Electronic Notes in Theoretical Computer Science*, *278*, 55-69. (Proceedings of the 7th Workshop on Methods for Modalities (M4M'2011))

Britz, K., Meyer, T., & Varzinczak, I. (2011b). Semantic foundation for preferential description logics. In D. Wang & M. Reynolds (Eds.), *Proceedings of the 24th australasian joint conference on artificial intelligence* (p. 491-500). Springer.

Britz, K., Meyer, T., & Varzinczak, I. (2012). Normal modal preferential consequence. In M. Thielscher & D. Zhang (Eds.), *Proceedings of the 25th australasian joint conference on artificial intelligence* (p. 505-516). Springer.

Britz, K., & Varzinczak, I. (n.d.). Preferential accessibility and preferred worlds. *Journal of Logic, Language and Information*. (To appear)

Britz, K., & Varzinczak, I. (2012). Defeasible modes of inference: A preferential perspective. In *Proceedings of the 14th international workshop on nonmonotonic reasoning (NMR).*

Britz, K., & Varzinczak, I. (2016a). Introducing role defeasibility in description logics. In L. Michael & A. Kakas (Eds.), *Proceedings of the 15th european conference on logics in artificial intelligence (JELIA)* (p. 174-189). Springer.

Britz, K., & Varzinczak, I. (2016b). Preferential modalities revisited. In *Proceedings of the 16th international workshop on nonmonotonic reasoning (NMR).*

Britz, K., & Varzinczak, I. (2017a). Context-based defeasible subsumption for $d\mathcal{SROIQ}$. In *Proceedings of the 13th International Symposium on Logical Formalizations of Commonsense Reasoning.*

Britz, K., & Varzinczak, I. (2017b). Toward defeasible $\mathcal{SROIQ}$. In *Proceedings of the 30th International Workshop on Description Logics.*

Casini, G., Meyer, T., Moodley, K., Sattler, U., & Varzinczak, I. (2015). Introducing defeasibility into OWL ontologies. In M. Arenas et al. (Eds.), *Proceedings of the 14th international semantic web conference (ISWC)* (p. 409-426). Springer.

Castilho, M., Gasquet, O., & Herzig, A. (1999). Formalizing action and change in modal logic I: the frame problem. *Journal of Logic and Computation*, *9*(5), 701-735.

Castilho, M., Herzig, A., & Varzinczak, I. (2002). It depends on the context! A decidable logic of actions and plans based on a ternary dependence relation. In *Proceedings of the 9th international workshop on nonmonotonic reasoning (NMR).*

Chellas, B. (1980). *Modal logic: An introduction.* Cambridge University Press.

Crocco, G., & Lamarre, P. (1992). On the connections between nonmonotonic inference systems and conditional logics. In R. Nebel, C. Rich, & W. Swartout (Eds.), *Proceedings of the 3rd international conference on principles of knowledge representation and reasoning (KR)* (p. 565-571). Morgan Kaufmann Publishers.

Delgrande, J. (1987). A first-order logic for prototypical properties. *Artificial Intelligence*, *33*, 105-130.

Delgrande, J. (1988). An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence*, *36*, 63-90.

Ditmarsch, H., Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic.* Springer.

Friedman, N., & Halpern, J. (2001). Plausibility measures and default reasoning. *Journal of the ACM*, *48*(4), 648-685.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*(6), 121-123.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2007). Preferential description logics. In N. Dershowitz & A. Voronkov (Eds.), *Logic for programming, artificial intelligence, and reasoning (LPAR)* (p. 257-272). Springer.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2008). Reasoning about typicality in preferential description logics. In S. Hölldobler, C. Lutz, & H. Wansing (Eds.), *Proceedings of the 11th european conference on logics in artificial intelligence (JELIA)* (p. 192-205). Springer.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2009a). Analytic tableaux calculi for KLM logics of nonmonotonic reasoning. *ACM Transactions on Computational Logic*, *10*(3), 18:1-18:47.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2009b). $\mathcal{ALC} + T$: a preferential extension of description logics. *Fundamenta Informaticae*, *96*(3), 341-372.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2012). A minimal model semantics for nonmonotonic reasoning. In L. Fariñas del Cerro, A. Herzig, & J. Mengin (Eds.), *Proceedings of the 13th european conference on logics in artificial intelligence (JELIA)* (p. 228-241). Springer.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2013). A non-monotonic description logic for reasoning about typicality. *Artificial Intelligence*, *195*, 165-202.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. (2015). Semantic characterization of rational closure: From propositional logic to description logics. *Artificial Intelligence*, *226*, 1-33.

Goré, R. (1999). Tableau methods for modal and temporal logics. In M. D'Agostino, D. Gabbay, R. Hähnle, & J. Posegga (Eds.), *Handbook of tableau methods* (p. 297-396). Kluwer Academic Publishers.

Governatori, G., Rotolo, A., & Calardo, E. (2012). Possible world semantics for defeasible deontic logic. In T. Agotnes, J. Broersen, & D. Elgesem (Eds.), *Proceedings of the 11th international conference on deontic logic in computer science (DEON)* (p. 44-60). Springer.

Halpern, J., & Moses, Y. (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, *54*, 319-379.

Hansson, B. (1969). An analysis of some deontic logics. *Noûs*, *3*, 373-398.

Karp, C. (1964). *Languages with expressions of infinite length.* North-Holland.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*, 167-207.

Ladner, R. (1977). The computational complexity of provability in systems of modal propositional logic. *SIAM Journal on Computing*, *6*(3), 467-480.

Laverny, N., & Lang, J. (2005). From knowledge-based programs to graded belief-based programs, part I: on-line reasoning. *Synthese*, *147*, 277-321.

Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, *55*, 1-60.

Lehrer, K., & Paxson, T. (1969). Undefeated justified true belief. *The Journal of Philosophy*, *66*(8), 225-237.

Lewis, D. (1973). *Counterfactuals.* Blackwell.

Lewis, D. (1974). Semantic analyses for dyadic deontic logic. In S. Stenlund (Ed.), *Logical theory and semantic analysis* (p. 1-14). D. Reidel Publishing Company.

Makinson, D. (1993). Five faces of minimality. *Studia Logica*, *52*, 339-379.

Shoham, Y. (1988). *Reasoning about change: Time and causation from the standpoint of artificial intelligence.* MIT Press.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (p. 98-112). Blackwell.

Varzinczak, I. (2002). *Causalidade e dependência em raciocínio sobre ações ("Causality and dependency in reasoning about actions").* M.Sc. thesis, Universidade Federal do Paraná, Curitiba, Brazil.

Veloso, P., Veloso, S., Viana, J., de Freitas, R., Benevides, M., & Delgado, C. (2009). On vague notions and modalities: a modular approach. *Logic Journal of the IGPL*, *18*(3), 381-402.

Appendix A. Proofs

A.1 *Proof of Proposition 1*

- Consistency: Assume $\mathscr{P} \Vdash \top \leadsto \bot$. Then $\min_\prec \llbracket \top \rrbracket^{\mathscr{P}} \subseteq \llbracket \bot \rrbracket^{\mathscr{P}}$, and therefore $\min_\prec \llbracket \top \rrbracket^{\mathscr{P}} \subseteq \emptyset$, from what follows $\min_\prec \llbracket \top \rrbracket^{\mathscr{P}} = \emptyset$, and therefore $W = \llbracket \top \rrbracket^{\mathscr{P}} = \emptyset$, which is a contradiction.
- Reflexivity: Straightforward, from the fact that $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \alpha \rrbracket^{\mathscr{P}}$.
- Left Logical Equivalence: Assume $\mathscr{P} \Vdash \alpha \leadsto \gamma$ and $\models \alpha \leftrightarrow \beta$. Then $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$. Since $\models \alpha \leftrightarrow \beta$, in particular we have $\llbracket \alpha \rrbracket^{\mathscr{P}} = \llbracket \beta \rrbracket^{\mathscr{P}}$, and therefore $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} = \min_\prec \llbracket \beta \rrbracket^{\mathscr{P}}$. Hence $\min_\prec \llbracket \beta \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$ and then $\mathscr{P} \Vdash \beta \leadsto \gamma$.
- And: Assume we have both $\mathscr{P} \Vdash \alpha \leadsto \beta$ and $\mathscr{P} \Vdash \alpha \leadsto \gamma$. Then $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \rrbracket^{\mathscr{P}}$ and $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$, and therefore $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \rrbracket^{\mathscr{P}} \cap \llbracket \gamma \rrbracket^{\mathscr{P}}$. Since $\llbracket \beta \rrbracket^{\mathscr{P}} \cap \llbracket \gamma \rrbracket^{\mathscr{P}} = \llbracket \beta \wedge \gamma \rrbracket^{\mathscr{P}}$, we have $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \wedge \gamma \rrbracket^{\mathscr{P}}$ and therefore $\mathscr{P} \Vdash \alpha \leadsto \beta \wedge \gamma$.
- Or: Assume we have both $\mathscr{P} \Vdash \alpha \leadsto \gamma$ and $\mathscr{P} \Vdash \beta \leadsto \gamma$. Let $w \in \min_\prec \llbracket \alpha \vee \beta \rrbracket^{\mathscr{P}}$. Then w is minimal in $\llbracket \alpha \rrbracket^{\mathscr{P}} \cup \llbracket \beta \rrbracket^{\mathscr{P}}$ and therefore $w \in \min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}}$ or $w \in \min_\prec \llbracket \beta \rrbracket^{\mathscr{P}}$. In either case, $w \in \llbracket \gamma \rrbracket^{\mathscr{P}}$. Hence $\mathscr{P} \Vdash \alpha \vee \beta \leadsto \gamma$.
- Right Weakening: Assume we have both $\mathscr{P} \Vdash \alpha \leadsto \beta$ and $\models \beta \rightarrow \gamma$. Then $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \rrbracket^{\mathscr{P}}$ and $\llbracket \beta \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$. From this follows $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$ and therefore $\mathscr{P} \Vdash \alpha \leadsto \gamma$.
- Cautious Monotony: Assume we have both $\mathscr{P} \Vdash \alpha \leadsto \beta$ and $\mathscr{P} \Vdash \alpha \leadsto \gamma$. Then $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \rrbracket^{\mathscr{P}}$ and $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \gamma \rrbracket^{\mathscr{P}}$. Let $w \in \min_\prec \llbracket \alpha \wedge \gamma \rrbracket^{\mathscr{P}}$. We show that $w \in \min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}}$. Suppose that this is not the case. Since $\prec$ is well founded, there must be $w' \in \min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}}$ such that $w' \prec w$. Because $\mathscr{P} \Vdash \alpha \leadsto \gamma$, we must also have $w' \in \llbracket \gamma \rrbracket^{\mathscr{P}}$, and therefore $w' \in \llbracket \alpha \rrbracket^{\mathscr{P}} \cap \llbracket \gamma \rrbracket^{\mathscr{P}}$, i.e., $w' \in \llbracket \alpha \wedge \gamma \rrbracket^{\mathscr{P}}$. From this and $w' \prec w$, it follows that w is not minimal in $\llbracket \alpha \wedge \gamma \rrbracket^{\mathscr{P}}$, which is a contradiction. Hence $w \in \min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}}$. From this and $\min_\prec \llbracket \alpha \rrbracket^{\mathscr{P}} \subseteq \llbracket \beta \rrbracket^{\mathscr{P}}$, it follows that $w \in \llbracket \beta \rrbracket^{\mathscr{P}}$ and therefore we have $\mathscr{P} \Vdash \alpha \wedge \gamma \leadsto \beta$. □

A.2 *Proof of Theorem 1*

We first show completeness of our tableau method, i.e., if $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ is preferentially valid, then every saturated tableau for $\neg\alpha$ is closed. Equivalently, if there is an open (saturated) tableau for α, then α is satisfiable, i.e., there exists a preferential Kripke model $\mathscr{P}$ in which $\llbracket \alpha \rrbracket^{\mathscr{P}} \neq \emptyset$.

In the following, we show that from any open tableau $\mathcal{T}$ for $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ one can construct a preferential Kripke model satisfying α, from which the result will then follow.

Let $\mathcal{T} := \mathcal{T}^\infty$ be an open saturated tableau for $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$. (Note that since the tableau rules are applied systematically, $\mathcal{T}^\infty$ is finite.) Then there must be an open branch $\langle \mathcal{S}, \Sigma, \prec \rangle$ in $\mathcal{T}$ (cf. Definition 13). Let the tuple $\mathscr{P} := \langle W_\mathcal{T}, R_\mathcal{T}, V_\mathcal{T}, \prec_\mathcal{T} \rangle$ be defined as follows:

- $W_\mathcal{T} := \{n \mid n :: \beta \in \mathcal{S}\}$;
- $R_\mathcal{T} := \langle R_1, \ldots, R_n \rangle$, where each $R_i := \Sigma(i)$, for $1 \leq i \leq n$;
- $V_\mathcal{T} := v$, where $v : W_\mathcal{T} \times \mathcal{P} \longrightarrow \{0, 1\}$ and $v(n, p) = 1$ if and only if $n :: p \in \mathcal{S}$;
- $\prec_\mathcal{T} := \prec^+$ (the transitive closure of $\prec$).

Lemma 1. *$\mathscr{P}$ is a preferential Kripke model.*

Proof. That $\mathscr{M}_{\mathcal{T}} := \langle W_{\mathcal{T}}, R_{\mathcal{T}}, V_{\mathcal{T}}\rangle$ is a Kripke model follows immediately from the definition of $W_{\mathcal{T}}$, $R_{\mathcal{T}}$ and $V_{\mathcal{T}}$ above. It remains to show that $\prec_{\mathcal{T}}$ is a smooth, strict partial order (Kraus et al., 1990), which amounts to showing that:

- $\prec_{\mathcal{T}}$ is irreflexive and transitive: This follows from the definition of $\prec_{\mathcal{T}}$ and the construction of $\prec$ in Rules $(\widetilde{\Diamond}_i)$, $(\Diamond_i)$, $(\leadsto)$ and $(\not\leadsto)$, since no pair (n, n) is ever added to $\prec$.
- $\prec_{\mathcal{T}}$ has no infinitely descending chains: Clearly, no pair (n, n') is added to $\prec$ beyond those added by Rules $(\widetilde{\Diamond}_i)$, $(\Diamond_i)$, $(\leadsto)$ and $(\not\leadsto)$. Given this, one can easily check that $\prec$ must have minimal elements.

□

It remains to show that $\mathscr{P}$ above satisfies α.

Lemma 2. *Let $\mathscr{P} = \langle W_{\mathcal{T}}, R_{\mathcal{T}}, V_{\mathcal{T}}, \prec_{\mathcal{T}}\rangle$ and let β be a sub-sentence of α. If $n :: \beta \in \mathcal{S}$, then $n \in [\![\beta]\!]^{\mathscr{P}}$.*

Proof. The proof is by structural induction on the number of connectives in β.

Base case: β is a literal. We have two cases: (i) $\beta = p \in \mathcal{P}$. Then $n :: p \in \mathcal{S}$ if and only if $v(n, p) = 1$ if and only if $V_{\mathcal{T}}(n, p) = 1$ if and only if $n \in [\![p]\!]^{\mathscr{P}} = [\![\beta]\!]^{\mathscr{P}}$. (ii) $\beta = \neg p$ for some $p \in \mathcal{P}$. Then $n :: \neg p \in \mathcal{S}$, and therefore $n :: p \notin \mathcal{S}$, otherwise $n :: \bot \in \mathcal{S}$ (as $\mathcal{T}$ is saturated), contradicting the assumption that $\langle \mathcal{S}, \Sigma, \prec\rangle$ is open. Hence $v(n, p) = 0$, and then $n \notin [\![p]\!]^{\mathscr{P}}$, from which follows $n \in W_{\mathcal{T}} \setminus [\![p]\!]^{\mathscr{P}} = [\![\neg p]\!]^{\mathscr{P}} = [\![\beta]\!]^{\mathscr{P}}$.

Induction step: The Boolean cases are as usual. We analyse the modal and conditional cases (below $\mathscr{M}_{\mathcal{T}} = \langle W_{\mathcal{T}}, R_{\mathcal{T}}, V_{\mathcal{T}}\rangle$):

- $\beta = \Box_i\gamma$: If $n :: \Box_i\gamma \in \mathcal{S}$, then $n' :: \gamma \in \mathcal{S}$ by Rule $(\Box_i)$, for every n' such that $(n, n') \in R_i$. By the induction hypothesis, $n' \in [\![\gamma]\!]^{\mathscr{P}}$, i.e., $\mathscr{M}_{\mathcal{T}}, n' \Vdash \gamma$ for every n' such that $(n, n') \in R_i$. From this we conclude $\mathscr{M}_{\mathcal{T}}, n \Vdash \Box_i\gamma$ and therefore $n \in [\![\Box_i\gamma]\!]^{\mathscr{P}}$.
- $\beta = \neg\Box_i\gamma$: If $n :: \neg\Box_i\gamma \in \mathcal{S}$, then by Rule $(\Diamond_i)$ there exists n' such that $(n, n') \in R_i$ and $n' :: \neg\gamma \in \mathcal{S}$. Then $n' \in [\![\neg\gamma]\!]^{\mathscr{P}}$, by the induction hypothesis. Hence $n \in [\![\neg\Box_i\gamma]\!]^{\mathscr{P}}$.
- $\beta = \widetilde{\Box}_i\gamma$: If $n :: \widetilde{\Box}_i\gamma \in \mathcal{S}$, then $n' :: \gamma \in \mathcal{S}$ by Rule $(\widetilde{\Box}_i)$, for every n' such that $n' \in \min_{\prec_{\mathcal{T}}} R_i(n)$. By the induction hypothesis, $n' \in [\![\gamma]\!]^{\mathscr{P}}$, and therefore $n \in [\![\widetilde{\Box}_i\gamma]\!]^{\mathscr{P}}$.
- $\beta = \neg\widetilde{\Box}_i\gamma$: If $n :: \neg\widetilde{\Box}_i\gamma \in \mathcal{S}$, then by Rule $(\widetilde{\Diamond}_i)$ there exists n' such that $n' \in \min_{\prec_{\mathcal{T}}} R_i(n)$ and $n' :: \neg\gamma \in \mathcal{S}$. Then $n' \in [\![\neg\gamma]\!]^{\mathscr{P}}$, by the induction hypothesis. Therefore we have $n \in [\![\neg\widetilde{\Box}_i\gamma]\!]^{\mathscr{P}}$.
- $\beta = \gamma \leadsto \gamma'$: If $n :: \gamma \leadsto \gamma' \in \mathcal{S}$, then by Rule $(\leadsto)$ we have either (i) $n :: \neg\gamma \in \mathcal{S}$; (ii) n is not minimal in $W^{\gamma}_{\mathcal{S}}$, or (iii) $n :: \gamma' \in \mathcal{S}$. By the induction hypothesis, from (i) and (iii) follows, respectively, $n \in [\![\neg\gamma]\!]^{\mathscr{P}}$ and $n \in [\![\gamma']\!]^{\mathscr{P}}$. In either case, $n \in [\![\gamma \leadsto \gamma']\!]^{\mathscr{P}}$. From (ii) and the induction hypothesis follows that n is not minimal in $[\![\gamma]\!]^{\mathscr{P}}$, hence we conclude $n \in [\![\gamma \leadsto \gamma']\!]^{\mathscr{P}}$.
- $\beta = \neg(\gamma \leadsto \gamma')$: If $n :: \neg(\gamma \leadsto \gamma') \in \mathcal{S}$, then by Rule $(\not\leadsto)$ we have that $n :: \gamma \in \mathcal{S}$, $n :: \neg\gamma' \in \mathcal{S}$, and n is set as minimal in $W^{\gamma}_{\mathcal{S}}$. By the induction hypothesis, $n \in [\![\gamma]\!]^{\mathscr{P}}$ and $n \notin [\![\gamma']\!]^{\mathscr{P}}$. Since n is minimal in $W^{\gamma}_{\mathcal{S}}$, it follows that $n \in [\![\neg(\gamma \leadsto \gamma')]\!]^{\mathscr{P}}$.

□

Now, since $0 :: \alpha \in \mathcal{S}$, from Lemma 2 we conclude that $0 \in [\![\alpha]\!]^{\mathscr{P}}$. Hence $[\![\alpha]\!]^{\mathscr{P}} \neq \emptyset$ for the preferential Kripke model constructed as above, and therefore α is satisfiable, as we wanted to show. □

In the following, we show soundness, i.e., if $\alpha \in \mathcal{L}^{\widetilde{\Box}+\leadsto}$ is (preferentially) satisfiable, then there is an open saturated tableau for α. Equivalently, if all the tableaux for α are closed, then α is unsatisfiable, i.e., $\neg\alpha$ is valid.

Definition 14. *Let $\mathcal{S}$ be a set of labeled sentences. $\mathcal{S}(n) := \{\beta \mid n :: \beta \in \mathcal{S}\}$.*

Definition 15. $\widehat{\mathcal{S}(n)} := \bigwedge\{\beta \mid \beta \in \mathcal{S}(n)\}$.

Lemma 3. *For every tableau rule in Figure 5, if for all possible $\mathcal{T}^{j+1} = \{\ldots, \langle\mathcal{S}^{j+1}, \Sigma^{j+1}, \prec^{j+1}\rangle, \ldots\}$ that can be obtained from $\mathcal{T}^j = \{\ldots, \langle\mathcal{S}^j, \Sigma^j, \prec^j\rangle, \ldots\}$[1] there is an n such that $\widehat{\mathcal{S}^{j+1}(n)}$ is (preferentially) unsatisfiable, then there is an n such that $\widehat{\mathcal{S}^j(n)}$ is (preferentially) unsatisfiable.*

Proof. We content ourselves with the cases of Rules $(\widetilde{\Diamond}_i)$, $(\Diamond_i)$, $(\leadsto)$ and $(\not\leadsto)$. (Rule $(\bot_\prec)$ is similar to Rule $(\bot)$.)

- Rule $(\widetilde{\Diamond}_i)$: If $\mathcal{S}^j$ contains $n :: \neg\widetilde{\Box}_i\beta$, then an application of Rule $(\widetilde{\Diamond}_i)$ creates a new label n', adds (n, n') to $\Sigma^j(i)$ to obtain $\Sigma^{j+1}(i)$, adds $n' :: \neg\beta$ to $\mathcal{S}^j$ to obtain $\mathcal{S}^{j+1}$, and sets n' as minimal in $\Sigma^{j+1}(i)$ with respect to $\prec^{j+1}$ (which extends $\prec^j$). Now, suppose there is n'' such that $\widehat{\mathcal{S}^{j+1}(n'')}$ is unsatisfiable. Then, either $n'' = n'$ or $n'' \neq n'$. If the latter is the case, then $\widehat{\mathcal{S}^j(n'')}$ is unsatisfiable (since $\widehat{\mathcal{S}^{j+1}(n)} = \widehat{\mathcal{S}^j(n)}$, for every $n \neq n'$), and therefore the lemma holds. If $n'' = n'$ is the case, then $\widehat{\mathcal{S}^{j+1}(n')}$ is unsatisfiable. But, since $\widehat{\mathcal{S}^{j+1}(n')} = \neg\beta$ (as $\mathcal{S}^{j+1} = \mathcal{S}^j \cup \{n' :: \neg\beta\}$ and n' is a freshly added label), then $\neg\beta$ must be unsatisfiable, i.e., $\models \beta$. From this and normal necessitation — Rule (RNN) —, we have $\models \widetilde{\Box}_i\beta$. Hence $\widehat{\mathcal{S}^j(n)}$ is unsatisfiable too because $n :: \neg\widetilde{\Box}_i\beta \in \mathcal{S}^j$.
- Rule $(\Diamond_i)$: If $\mathcal{S}^j$ contains $n :: \neg\Box_i\beta$, then an application of Rule $(\Diamond_i)$ will create a new label n' and either
  (1) Add (n, n') to $\Sigma^j(i)$ to obtain $\Sigma^{j+1}_{(1)}(i)$, add $n' :: \neg\beta$ to $\mathcal{S}^j$ to obtain $\mathcal{S}^{j+1}_{(1)}$, and set n' as *minimal* in $\Sigma^{j+1}_{(1)}(i)$ w.r.t. $\prec^{j+1}_{(1)}$ (thereby extending $\prec^j$), or
  (2) Add (n, n') to $\Sigma^j(i)$ to obtain $\Sigma^{j+1}_{(2)}(i)$, add $n' :: \neg\beta$ to $\mathcal{S}^j$ to obtain $\mathcal{S}^{j+1}_{(2)}$, create a new label n'' and also add (n, n'') to $\Sigma^{j+1}_{(2)}(i)$, add (n'', n') to $\prec^j$ to obtain $\prec^{j+1}_{(2)}$ and set n'' as minimal in $\Sigma^{j+1}_{(2)}(i)$ w.r.t. $\prec^{j+1}_{(2)}$, in other words, set n' as *not minimal* in $\Sigma^{j+1}_{(2)}(i)$ w.r.t. $\prec^{j+1}_{(2)}$.

  Now suppose there are n_1 and n_2 such that both $\widehat{\mathcal{S}^{j+1}_{(1)}(n_1)}$ and $\widehat{\mathcal{S}^{j+1}_{(2)}(n_2)}$ are unsatisfiable (i.e., both branches explored after the rule application are closed). If $n_1 = n$ or $n_2 = n$ or both, then the lemma holds for the same reason given in the case of Rule $(\widetilde{\Diamond}_i)$ above. If neither $n_1 = n$ nor $n_2 = n$, then we must have $n_1 = n_2 = n'$ (the unsatisfiability of each of $\mathcal{S}^{j+1}_{(1)}$ and $\mathcal{S}^{j+1}_{(2)}$ is brought in by the new added world and its minimality or not w.r.t. $\prec^{j+1}_{(1)}$ and $\prec^{j+1}_{(2)}$). From this, and looking again at cases (1) and (2) above, we conclude that $\widehat{\mathcal{S}^{j+1}_{(1)}(n')}$ and $\widehat{\mathcal{S}^{j+1}_{(2)}(n')}$ are both unsatisfiable if and only if
  - $\neg\beta$ is unsatisfiable or n' is not minimal, and
  - $\neg\beta$ is unsatisfiable or n' is minimal

[1] In fact, there is only one possibility, except in the cases of Rules $(\vee)$, $(\Diamond_i)$ and $(\leadsto)$.

from which it follows that either $\neg\beta$ is unsatisfiable or n' *is and is not* minimal. Therefore, $\neg\beta$ is unsatisfiable, i.e., $\models \beta$. From this and the necessitation rule, we have $\models \Box_i\beta$. Hence $\widehat{\mathcal{S}^j(n)}$ is unsatisfiable too because $n :: \neg\Box_i\beta \in \mathcal{S}^j$.

- Rule ($\rightsquigarrow$): If $\mathcal{S}^j$ contains $n :: \alpha \rightsquigarrow \beta$, then an application of Rule ($\rightsquigarrow$) will either
  (1) Add $n :: \neg\alpha$ to $\mathcal{S}^j$ to obtain $\mathcal{S}^{j+1}_{(1)}$, or
  (2) Create a new label n', add $n' :: \alpha$ to obtain $\mathcal{S}^{j+1}_{(2)}$ and set $n' \prec n$ (thereby extending $\prec^j$), or
  (3) Add $n :: \beta$ to obtain $\mathcal{S}^{j+1}_{(3)}$.

  Now suppose there are n_1, n_2 and n_3 such that each of $\widehat{\mathcal{S}^{j+1}_{(1)}(n_1)}$, $\widehat{\mathcal{S}^{j+1}_{(2)}(n_2)}$ and $\widehat{\mathcal{S}^{j+1}_{(3)}(n_3)}$ is unsatisfiable (i.e., all branches explored after the rule application are closed). If $n_1 \neq n$ or $n_3 \neq n$ or both, then either $\widehat{\mathcal{S}^j(n_1)}$ or $\widehat{\mathcal{S}^j(n_3)}$ is unsatisfiable, since $\mathcal{S}^{j+1}_{(1)} \setminus \mathcal{S}^j = \{n :: \neg\alpha\}$ and $\mathcal{S}^{j+1}_{(3)} \setminus \mathcal{S}^j = \{n :: \beta\}$. If $n_2 = n$, the lemma also follows. Hence, there remains one case to be explored, namely $n_1 = n_3 = n$ and $n_2 = n'$. From this, and looking again at cases (1)–(3) above, we conclude that $\widehat{\mathcal{S}^{j+1}_{(1)}(n)}$, $\widehat{\mathcal{S}^{j+1}_{(2)}(n')}$ and $\widehat{\mathcal{S}^{j+1}_{(3)}(n)}$ are all unsatisfiable if and only if

  - $\widehat{\mathcal{S}^j(n)} \models \alpha \wedge \neg\beta$, and
  - Either $\models \neg\alpha$ (because $\mathcal{S}^{j+1}_{(2)}(n') = \{\alpha\}$ and $\widehat{\mathcal{S}^{j+1}_{(2)}(n')}$ is unsatisfiable) or there can be no label more preferred than n w.r.t. $\prec^{j+1}$.

  Now, if $\models \neg\alpha$, then $\widehat{\mathcal{S}^j(n)} \models \bot$ and the lemma follows. If $\not\models \neg\alpha$, then, since $\widehat{\mathcal{S}^j(n)} \models \alpha$, we must have n minimal in $W^{\alpha}_{\mathcal{S}}$ w.r.t. $\prec^{j+1}$. It is not hard to see that there can be no preferential Kripke model with a possible world satisfying all sentences in $\mathcal{S}^j(n)$ and that also satisfies this minimality constraint, for, if there were such a possible world in a preferential model, then it would satisfy $\alpha \rightsquigarrow \beta$ and $\neg(\alpha \rightsquigarrow \beta)$, which is absurd. Hence $\widehat{\mathcal{S}^j(n)}$ is preferentially unsatisfiable.

- Rule ($\not\rightsquigarrow$): If $\mathcal{S}^j$ contains $n :: \neg(\alpha \rightsquigarrow \beta)$, then an application of Rule ($\not\rightsquigarrow$) adds $n :: \alpha$ and $n :: \neg\beta$ to $\mathcal{S}^j$ to obtain $\mathcal{S}^{j+1}$, and sets n as minimal in $W^{\alpha}_{\mathcal{S}}$ w.r.t. $\prec^{j+1}$ (which extends $\prec^j$). Now, suppose $\widehat{\mathcal{S}^{j+1}(n')}$ is unsatisfiable for some n'. Then, $\widehat{\mathcal{S}^j(n')} \models \alpha \rightsquigarrow \beta$. If $n' \neq n$, then $\mathcal{S}^{j+1}(n') = \mathcal{S}^j(n')$ (because the rule application concerns only sentences labeled with n), from which the lemma follows. If $n' = n$, then $\widehat{\mathcal{S}^j(n)}$ is unsatisfiable, since $n :: \neg(\alpha \rightsquigarrow \beta) \in \mathcal{S}^j$.

□

From Lemma 3 we conclude that if all tableaux for α are closed, then every $\widehat{\mathcal{S}(n)}$ is unsatisfiable. In particular $\widehat{\mathcal{S}(0)} = \alpha$ is unsatisfiable, too. Hence all the rules preserve satisfiability when transforming one set of branches into another one. This warrants soundness of our tableau rules. □

By putting the above results together, we get to the proof of the theorem.

Theorem 1. *The tableau calculus for $\mathcal{L}^{\tilde{\Box}+\rightsquigarrow}$ is sound and complete with respect to our modal preferential semantics.*

Proof. Soundness follows from Lemma 3. Completeness is established by Lemmas 1 and 2. □