Noname manuscript No.
(will be inserted by the editor)

Preferential Accessibility and Preferred Worlds

Katarina Britz · Ivan Varzinczak

Received: date / Accepted: date

Abstract Modal accounts of normality in non-monotonic reasoning traditionally have an underlying semantics based on a notion of preference amongst worlds. In this paper, we motivate and investigate an alternative semantics, based on ordered accessibility relations in Kripke frames. The underlying intuition is that some world tuples may be seen as more normal, while others may be seen as more exceptional. We show that this delivers an elegant and intuitive semantic construction, which gives a new perspective on defeasible necessity. Technically, the revisited logic does not change the expressive power of our previously defined preferential modalities. This conclusion follows from an analysis of both semantic constructions via a generalisation of bisimulations to the preferential case. Reasoners based on the previous semantics therefore also suffice for reasoning over the new semantics. We complete the picture by investigating different notions of defeasible conditionals in modal logic that can also be captured within our framework.[1]

Keywords Modal logic, non-monotonic reasoning, preferential semantics

1 Introduction and Motivation

Accounts of normality, typicality, plausibility and alike traditionally have an underlying semantics built on a notion of preference on *worlds*. Such is the

K. Britz
CSIR-SU CAIR, Stellenbosch University, South Africa
E-mail: abritz@sun.ac.za

I. Varzinczak
CRIL, Univ. Artois & CNRS, France
E-mail: varzinczak@cril.fr

[1] A preliminary version of the work reported in this paper was presented at the Workshop on Nonmonotonic Reasoning [19].

case of non-monotonic entailment [36,45,46], conditionals [10,38], belief revision [3,4,35], counterfactuals [39,47], obligations [31,40,44] and many others, as known from the literature on non-monotonic reasoning, conditional and deontic logics, and related areas. Roughly speaking, the usual approach consists in selecting some possible worlds (or propositional valuations) as being more normal, and carrying out the reasoning relative to an underlying normality ordering on worlds.

A typical representative of these threads of investigation is the well-known preferential approach [46] and its derivatives [36,38]. There, a preference relation is defined on the set of possible worlds with the tacit assumption that these suffice to reason about what is normal or expected. A case can indeed be made for such an assumption in a propositional setting. However, in logics with more structure, it is reasonable to say that the focus should not be confined to the relative normality of worlds, but rather be on whatever structure we have at our disposal in the respective underlying semantics. To witness, in a modal-logic context, it makes sense to ask whether some links between worlds in a frame are more normal than, or preferred to others, irrespective of whether the worlds involved are somehow comparable amongst themselves. In other words, one can be interested in the normality of the *transition* from one world to another one. This point is better illustrated with some well-known concrete applications of modal logics as given below.

Let us assume a very simple scenario in which we have only one propositional atom, on, of which the intuition is that a particular light-bulb is on. Moreover, let us assume there is only one action at one's disposal, namely toggle (hereafter abbreviated t), of which the intuition is that of changing the state of the light switch. Figure 1 below depicts a possible-worlds model for this scenario.

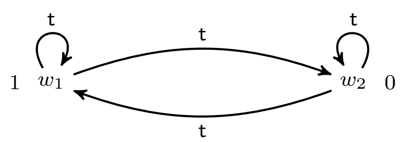

Fig. 1 A possible-worlds model for one action and one atom.

Intuitively, normal executions of the toggle action are given by the t-transitions from w_1 to w_2 and back, whereas the reflexive arrows are exceptional in the given scenario. Therefore, it becomes important to single out those executions of the action that are deemed normal from those that are exceptional. In Figure 1, this would amount to enriching the semantic structure (in a way still to be defined) with information specifying that the pairs (w_1, w_2) and (w_2, w_1) take precedence over (w_1, w_1) and (w_2, w_2) when reasoning about possible executions of the action.

Let us now consider an epistemic variant of the above scenario, in which we have only one atomic proposition carbon, the intuition of which is that a specific model car meets EU standard for carbon emissions, as measured by an approved emissions test and checked by an agent A with associated epistemic accessibility relation. Figure 2 below depicts one possible configuration of such a scenario.

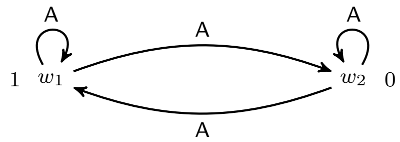

Fig. 2 A possible-worlds model for one agent and one atom.

Since the emissions test is a reliable indicator, agent A knows whether a given car meets EU emissions standard. However, the agent admits the (outrageous) possibility of tampering with the test. In this case, we could say that A *defeasibly knows* whether a car meets the EU emissions standard. The epistemic stance adopted by the agent is to prefer (w_1, w_1) and (w_2, w_2) in Figure 2, which, in this example, are more normal than (w_1, w_2) and (w_2, w_1). It is not hard to see that a similar motivation above also holds in a doxastic context, as certain beliefs may be more entrenched than others.

In this example, as in the previous one, we could explicitly model exceptionality by introducing further atomic propositions to the language, but that would defeat the objective of defining a semantics for reasoning about exceptionality, and further demand prior knowledge about different degrees of exceptionality.

In order to motivate the foregoing ideas in a deontic context, let us assume a language with a single propositional atom, namely fair-play, henceforth abbreviated f, of which the intuition is that, in a competition, the players abide by an established standard of 'honorable conduct'. In this context, adopting a fair-play stance is not to be seen as an obligation in the usual (strict) meaning of the term. It is rather a matter of best practice in that it corresponds to the expected, though not enforceable (even if, in some cases, liability-biding), attitude. Figure 3 below depicts a possible-worlds model for this scenario.

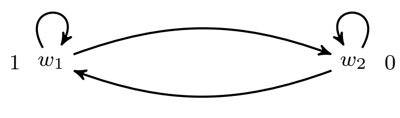

Fig. 3 A possible-worlds model for one atom in a deontic-context.

Then an f-world is a preferred alternative to a $\neg\mathsf{f}$-world. Semantically, this requirement can be translated as setting the pairs (w_1, w_1) and (w_2, w_1) as more preferred than (w_2, w_2) and (w_1, w_2). In this specific example, it happens that we could also model the underlying preference as an ordering on worlds, with f-worlds preferred to the $\neg\mathsf{f}$-worlds. However, in the preceding examples, ordering worlds rather than pairs of worlds is neither intuitive nor is it immediately clear whether this is even possible.

In this paper, we present an alternative semantics for preferential modalities, obtained by shifting the normality spotlight from possible worlds to transitions amongst them, i.e., to accessibility relations in Kripke frames. The justification for doing so stems from a comparison with the classical (monotonic) case: In classical Kripke semantics, modalities are primarily about accessibility, and only secondarily about worlds. Hence, we contend that accounts of a notion of defeasibility in modalities (like those illustrated above) should primarily focus on normality of the accessibility relations rather than, or at least prior to, that of the (accessible) worlds. With that we hope to pave the way for further explorations of non-monotonicity in modal logics, in particular in extensions of the preferential approach [13].

The remainder of the present paper is structured as follows: After some preliminaries on both basic modal logics and its preferential extension we have defined in previous work (Section 2), we introduce an alternative preferential semantics in which the notion of normality applies to transitions in a Kripke frame rather than to possible worlds, and we use it as the basis for a new logic of defeasible modalities (Section 3). In Section 4, we extend the standard notion of bisimulation to the preferential case and in Section 5 we apply it in comparing the expressive power of our new account of modal defeasibility with the one from our previous work. In Section 6, we explore how to enrich our language with defeasible conditionals, in particular how these can be given a semantics in terms of our new preferential structures. Before concluding, we address related work and comment on open questions we leave for future investigation (Section 7).

2 Background

In this section, we provide the required formal background for the rest of this work. In particular, we set up the notation and conventions that shall be followed in the upcoming sections. (The reader conversant with modal logic may skip to Section 2.2.)

2.1 Modal Logic

We assume a set $\mathcal{P}$ of *atomic propositions*, denoted $p, q, \ldots$. Sentences are denoted by $\alpha, \beta, \ldots$ and are constructed as follows $(1 \leq i \leq n)$:

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

All the other truth-functional connectives are defined in terms of $\neg$ and $\wedge$ in the usual way. Given $\Box_i$, $\Diamond_i\alpha := \neg\Box_i\neg\alpha$. We use $\top$ and $\bot$ as abbreviations for, respectively, $p \vee \neg p$ and $p \wedge \neg p$, for $p \in \mathcal{P}$.

With $\mathcal{L}^{\Box}$ we denote the *language* of all modal sentences. 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 $V\colon W \longrightarrow \{0,1\}^{\mathcal{P}}$ is a valuation function mapping possible worlds into propositional valuations.

As an example, Figure 4 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 $R_a := \{(w_1, w_2), (w_1, w_3), (w_4, w_3)\}$, and $R_b := \{(w_1, w_4), (w_2, w_3)\}$, and V_1 as in Figure 4.

In our pictorial representations of models, we represent 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.

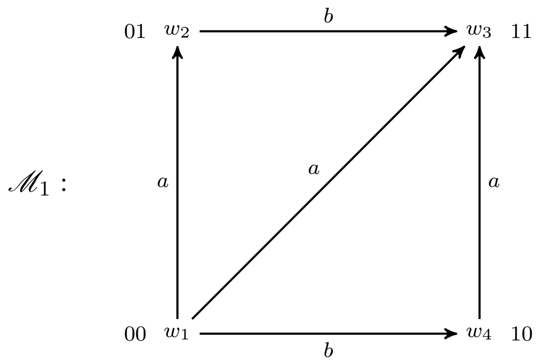

Fig. 4 A Kripke model for $\mathcal{P} = \{p, q\}$ and two modalities, a and b.

We shall use $w, u, v, \ldots$ (possibly decorated with primes) to denote possible worlds. Moreover, where it aids readability, we shall henceforth sometimes write tuples of the form (w, w') as ww'.

Sentences of $\mathcal{L}^{\Box}$ are true or false relative to a possible world in a given Kripke model:

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

- Propositional sentences are evaluated as usual;
- $\mathscr{M}, w \Vdash \Box_i\alpha$ iff $\mathscr{M}, w' \Vdash \alpha$ 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 $w \in W$ such that $\mathscr{M}, w \Vdash \alpha$. α is *valid* if $\mathscr{M}, w \Vdash \alpha$ for every w in every $\mathscr{M}$.

Given $\mathcal{K} \subseteq \mathcal{L}^{\Box}$ and $\alpha \in \mathcal{L}^{\Box}$, we say that $\mathcal{K}$ *entails* α, denoted $\mathcal{K} \models \alpha$, if and only if for every Kripke model $\mathscr{M}$ and every w in $\mathscr{M}$, if $\mathscr{M}, w \Vdash \beta$ for every $\beta \in \mathcal{K}$, then $\mathscr{M}, w \Vdash \alpha$.

For more details on modal logic, we refer the reader to the handbook by Blackburn et al. [6]. Our adoption of local entailment is also in line with the modal tradition followed there.

2.2 Preferential Modalities

In previous work [13,15,17,16], we enriched the standard Kripke semantics with a preference relation on the set of possible worlds. The underlying motivation for the resulting semantic structure is similar to that of Boutilier's CT4O models [10], the plausibility models of Baltag and Smets [3,4], and Giordano et al.'s preferential DL interpretations [28].

Definition 3 (W-Ordered Model) A *W-ordered model* is a tuple $\mathscr{W} := \langle W, R, V, \prec \rangle$ where $\langle W, R, V \rangle$ is as in Definition 1 and $\prec \subseteq W \times W$ is a well-founded strict partial order on W, i.e., $\prec$ is irreflexive, transitive and every non-empty $X \subseteq W$ has minimal elements w.r.t. $\prec$. For $X \subseteq W$, $\min_{\prec} X := \{w \in X \mid \text{there is no } w' \in X \text{ such that } w' \prec w\}$ denotes the *minimal elements* of X with respect to the preference relation $\prec$.

The intuition behind the preference relation $\prec$ in a W-ordered model $\mathscr{W}$ is that the worlds lower down in the ordering are deemed as more preferred (or more normal) than those higher up.

As an example, the W-ordered model $\mathscr{W}_1 = \langle W_1, R_1, V_1, \prec_1 \rangle$ is depicted in Figure 5 below, where $\langle W_1, R_1, V_1 \rangle$ is as in Figure 4 and $\prec_1 := \{(w_1, w_2), (w_2, w_3), (w_1, w_3), (w_4, w_3)\}$.

We can then extend $\mathcal{L}^{\Box}$ with a family of defeasible modal operators $\mathrel{\approx\!\!\!\Box}_i$ (called 'flag'), $1 \leq i \leq n$, where n is the number of classical modalities in the language. The sentences of the extended language are then recursively defined by the following grammar ($1 \leq i \leq n$):

$$\alpha ::= p \mid \neg\alpha \mid (\alpha \wedge \alpha) \mid \Box_i \alpha \mid \mathrel{\approx\!\!\!\Box}_i \alpha$$

As before, the other connectives are defined in terms of $\neg$ and $\wedge$ in the usual way, $\top$ and $\bot$ are seen as abbreviations, and $\Diamond_i$ is the dual of $\Box_i$. Moreover, with $\mathrel{\approx\!\!\!\Diamond}_i$ (called 'flame') we denote the dual of $\mathrel{\approx\!\!\!\Box}_i$. With $\mathcal{L}^{\mathrel{\approx\!\!\!\Box}}$ we denote the set of all sentences of such a richer language.

Definition 4 (Truth Conditions for $\mathcal{L}^{\mathrel{\approx\!\!\!\Box}}$) Let $\mathscr{W} = \langle W, R, V, \prec \rangle$ be a W-ordered model, and let $w \in W$.

- $\mathcal{L}^{\Box}$-sentences are evaluated as usual (Definition 2);
- $\mathscr{W}, w \Vdash \mathrel{\approx\!\!\!\Box}_i \alpha$ iff $\mathscr{W}, w' \Vdash \alpha$ for all $w' \in \min_{\prec} R_i(w)$.

$\mathscr{W}_1$:

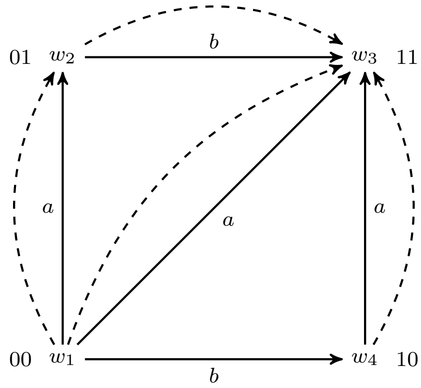

Fig. 5 A W-ordered model for $\mathcal{P} = \{p, q\}$ and two modalities (a and b). The preference relation $\prec_1$ is represented by the dashed arrows, which point from more preferred to less preferred worlds.

The notions of satisfaction, validity and entailment are generalised to $\mathcal{L}^{\widetilde{\Box}}$-sentences and W-ordered models in the obvious way.

Informally, a sentence of the form $\widetilde{\Box}_i\alpha$ holds in a world if α holds in all the most preferred amongst its i-successors. It is easy to see that $\widetilde{\Box}$ is weaker than $\Box$, i.e., the following is a validity ($1 \leq i \leq n$):

$$\models \Box_i\alpha \rightarrow \widetilde{\Box}_i\alpha$$

Hence, intuitively, flag can be read as *defeasible necessity.*

As an example, considering the W-ordered model $\mathscr{W}_1$ from Figure 5, we have that $\mathscr{W}_1, w_1 \Vdash \widetilde{\Box}_a\neg p$ (but note that $\mathscr{W}_1, w_1 \nVdash \Box_a\neg p$).

Finally, a sound, complete and terminating tableau-based proof procedure for $\mathcal{L}^{\widetilde{\Box}}$ has been defined [16,17], of which the complexity of the satisfiability problem has been shown to be PSPACE-complete, hence the same complexity class as the underlying modal logic we started off with.

3 Preferential Modalities Revisited

In spite of the gain in expressiveness when checked against traditional approaches to defeasible reasoning [16,17], $\widetilde{\Box}$does not allow us to model the type of reasoning motivated in the Introduction inasmuch as it relies on orderings on worlds. In this section, we revisit the framework for preferential modalities, in particular its semantic constructions.

3.1 R-Ordered Models

We start by giving a formal account of the semantic ideas put forward in the Introduction.

Definition 5 (R-Ordered Model) An R-*ordered model* is a tuple $\mathscr{R} := \langle W, R, V, \ll \rangle$ where W is a (non-empty and possibly infinite) 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, for $1 \leq i \leq n$, $V \colon W \longrightarrow \{0,1\}^{\mathcal{P}}$ is a valuation function assigning each world to a valuation on $\mathcal{P}$, and $\ll := \langle \ll_1, \ldots, \ll_n \rangle$, where each $\ll_i \subseteq R_i \times R_i$, for $1 \leq i \leq n$, is a well-founded strict partial order on the respective R_i, i.e., each $\ll_i$ is irreflexive, transitive and every non-empty $X \subseteq R_i$ has minimal elements w.r.t. $\ll_i$. For $X \subseteq R_i$, $\min_{\ll_i} X := \{(w, w') \in X \mid$ there is no $(u, v) \in X$ such that $(u, v) \ll_i (w, w')\}$ denotes the *minimal elements* of X with respect to the preference relation $\ll_i$ associated to R_i.

Given $\mathscr{R} := \langle W, R, V, \ll \rangle$, the intuition of W, R and V is the same as that in a standard Kripke model (Definition 1). The intuition of each $\ll_i$ in $\ll$ is that the pairs (w, w') lower down in the ordering $\ll_i$ are deemed as the most normal (or typical, or expected) in the context of R_i.

Since we assume each $\ll_i$ to be a well-founded strict partial order on the respective R_i, we are guaranteed that for every $X \subseteq R_i$ such that $X \neq \emptyset$, $\min_{\ll_i} X$ is well defined.

As an example, the R-ordered model $\mathscr{R}_1 := \langle W_1, R_1, V_1, \ll_1 \rangle$ is depicted in Figure 6, where $\langle W_1, R_1, V_1 \rangle$ is as in Figure 4, and $\ll_1 := \langle \ll_a, \ll_b \rangle$, with $\ll_a := \{(w_1w_2, w_1w_3), (w_1w_3, w_4w_3), (w_1w_2, w_4w_3)\}$ and $\ll_b := \{(w_1w_4, w_2w_3)\}$, represented, respectively, by the dashed and the dotted arrows in the picture. (Note the direction of the $\ll$-arrows, which point from more preferred to less preferred transitions.) For the sake of readability, in our pictorial representations of R-ordered models, we shall omit the transitive $\ll$-arrows.

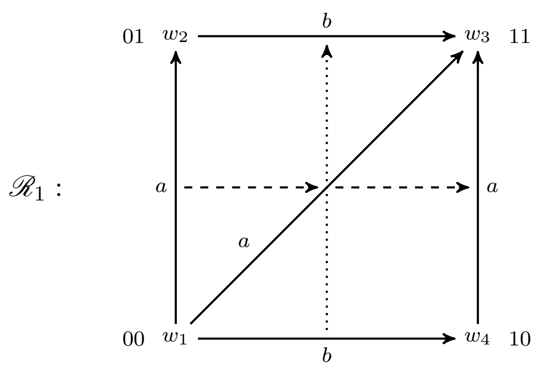

Fig. 6 An R-ordered model for $\mathcal{P} = \{p, q\}$ and two modalities. The preference relation $\ll_a$ is represented by the dashed arrows, whereas $\ll_b$ by the dotted one.

Satisfaction of $\mathcal{L}^{\Box}$-sentences by R-ordered models is defined in the obvious way. It is easy to see that the addition of the $\ll$-component to Kripke models preserves the truth of all $\mathcal{L}^{\Box}$-sentences holding in the remaining structure:

Observation 1 *Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ be an R-ordered model and let $\mathscr{R}_{\mathscr{M}} = \langle W, R, V\rangle$. For every $\alpha \in \mathcal{L}^{\Box}$, $\mathscr{R} \Vdash \alpha$ iff $\mathscr{R}_{\mathscr{M}} \Vdash \alpha$.*

3.2 A New Logic of Defeasible Modalities

We shall now enrich our underlying modal language with a family of additional modal operators $\approxeq_i$, $1 \leq i \leq n$, where n is the number of classical modalities in the language. (We call $\approxeq$ the 'banner'.) The sentences of the extended modal language are recursively defined as follows ($1 \leq i \leq n$):

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

With $\mathcal{L}^{\approxeq}$ we denote the set of all sentences of the banner language.

Definition 6 Let $\mathscr{R} = \langle W, R, V, \ll\rangle$. For every $w \in W$ and every $R_i \subseteq W \times W$, we define:

$$R_i^w := \{(u, v) \mid (u, v) \in R_i \text{ and } u = w\}$$

Definition 7 ($\mathcal{L}^{\approxeq}$ Truth Conditions) Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ be an R-ordered model and $w \in W$.

- $\mathcal{L}^{\Box}$-sentences are evaluated as usual;
- $\mathscr{R}, w \Vdash \approxeq_i\alpha$ iff $\mathscr{R}, w' \Vdash \alpha$ for every w' such that $(w, w') \in \min_{\ll_i} R_i^w$.

The notions of satisfaction, validity and entailment are generalised to $\mathcal{L}^{\approxeq}$-sentences and R-ordered models in the obvious way.

Informally, a sentence of the form $\approxeq_i\alpha$ holds in a world if α holds in all its *most normally i-accessible worlds*. (Note that *normally accessible* differs from *normal amongst the accessible*—cf. Definition 4.) As an example, in the R-ordered model $\mathscr{R}_1$ of Figure 6, we have that $\mathscr{R}_1, w_1 \Vdash \approxeq_a\neg p$ (but, of course, $\mathscr{R}_1, w_1 \not\Vdash \Box_a\neg p$).

For every $\alpha \in \mathcal{L}^{\approxeq}$, let $[\![\alpha]\!]^{\mathscr{R}}$ denote the α-worlds in $\mathscr{R}$. We then have the following result:

Proposition 1 *The operator $\approxeq$ satisfies the following properties for every $\alpha \in \mathcal{L}^{\approxeq}$ and every $\mathscr{R}$:*

- *It is ampliative w.r.t. $\Box$, i.e., $[\![\Box\alpha]\!]^{\mathscr{R}} \subseteq [\![\approxeq\alpha]\!]^{\mathscr{R}}$;*
- *It is monotone (increasing) on $\mathscr{P}(W)$, i.e., if $[\![\alpha]\!]^{\mathscr{R}} \subseteq [\![\beta]\!]^{\mathscr{R}}$, then also $[\![\approxeq\alpha]\!]^{\mathscr{R}} \subseteq [\![\approxeq\beta]\!]^{\mathscr{R}}$;*

Hence, $≋$ provides an alternative perspective on the notion of defeasible necessity as formalised by $\approx\!\!\!\!\!\Box$. For instance, in an action context, some executions (which refer to transitions) of a given action are deemed as more normal than others. A priori, this is different from saying that some effects (which refer to target worlds) are normal. Indeed, an abnormal execution may still lead to the expected (normal) effect, just as a normal execution may produce an abnormal effect. (We shall come back to this issue later on.)

An immediate consequence of ampliativity is that $≋$ is weaker than $\Box$:

$$\models \Box_i\alpha \rightarrow ≋_i\alpha, \text{ for } 1 \leq i \leq n.$$

Speaking of validities, here we shall not provide an axiomatisation for $≋$ in the customary way. This is on purpose, as the results of subsequent sections will shed light on a proof system for $\mathcal{L}^{≋}$. For now we shall content ourselves with the following result:

Theorem 1 *$\mathcal{L}^{≋}$ has the finite-model property.*

Proof See Appendix A ■

The definitions of R-ordered models and $≋$ raise the question as to how $\mathcal{L}^{≋}$ and $\mathcal{L}^{\approx\!\!\!\!\!\Box}$ compare to each other in terms of expressive power. Before we answer this question, in the next section we generalise the notion of bisimulations to the preferential case.

4 Preferential Bisimulations

Standard bisimulations are used to determine whether two Kripke models have the same modal properties, and to reason about modal expressivity. Here, we extend the definition of bisimulations to W-ordered and R-ordered models so that in the next section we can use it to make precise the connection between these notions, and the resulting modalities and modal languages.

Definition 8 [6] Let $\mathscr{M} = \langle W, R, V\rangle$ and $\mathscr{M}' = \langle W', R', V'\rangle$. A *bisimulation* between $\mathscr{M}$ and $\mathscr{M}'$ is a non-empty binary relation E between their domains (that is, $\mathsf{E} \subseteq W \times W'$) such that, whenever $w\mathsf{E}w'$, we have that:

1. For every $p \in \mathcal{P}$, $\mathscr{M}, w \Vdash p$ if and only if $\mathscr{M}', w' \Vdash p$;
2. if wR_iv, then there is v' in W' such that $v\mathsf{E}v'$ and $w'R'_iv'$, and
3. if $w'R'_iv'$, then there is v in W such that $v\mathsf{E}v'$ and wR_iv.

Informally, two worlds are bisimilar if they satisfy the same atomic information and their accessibility structures match. Two pointed models $(\mathscr{M}, w)$ and $(\mathscr{M}', w')$ are bisimilar if there is a bisimulation E between $\mathscr{M}$ and $\mathscr{M}'$ such that $w\mathsf{E}w'$. It then follows that:

Lemma 1 (Bisimulation invariance lemma [6]) *If E is a bisimulation between $\mathscr{M} = \langle W, R, V\rangle$ and $\mathscr{M}' = \langle W', R', V'\rangle$, $w \in W$ and $w' \in W'$, and $w\mathsf{E}w'$, then w and w' satisfy the same basic modal sentences.*

The next definition and lemma generalise bisimulations to take account of a preference order on worlds, as defined on models of $\mathcal{L}^{\precsim}$. Informally, two worlds are bisimilar if they satisfy the same atomic information and their modal accessibility structures match, both with respect to accessible worlds and with respect to most preferred relative accessible worlds. Bisimilar worlds then also satisfy the same preferential modal sentences.

Definition 9 **(*W*-ordered bisimulation)** Let $\mathscr{W} = \langle W, R, V, \prec\rangle$ and $\mathscr{W}' = \langle W', R', V', \prec'\rangle$ be *W*-ordered models. A *W-ordered bisimulation* between $\mathscr{W}$ and $\mathscr{W}'$ is a non-empty binary relation $\mathsf{E} \subseteq W \times W'$ such that, whenever $w\mathsf{E}w'$, we have that:

1. For every $p \in \mathcal{P}$, $\mathscr{W}, w \Vdash p$ if and only if $\mathscr{W}', w' \Vdash p$;
2. if wR_iv, then there is v' in W' such that $v\mathsf{E}v'$ and $w'R'_iv'$, and
   - if $v \in \min_{\prec} R_i(w)$, then $v' \in \min_{\prec'} R'_i(w')$;
3. if $w'R'_iv'$, then there is v in W such that $v\mathsf{E}v'$ and wR_iv, and
   - if $v' \in \min_{\prec'} R'_i(w')$, then $v \in \min_{\prec} R_i(w)$.

Lemma 2 **(*W*-ordered bisimulation invariance lemma)** *If* E *is a W-ordered bisimulation between* $\mathscr{W} = \langle W, R, V, \prec\rangle$ *and* $\mathscr{W}' = \langle W', R', V', \prec'\rangle$*, and* $w\mathsf{E}w'$*, then* w *and* w' *satisfy the same modal sentences in the extended modal language* $\mathcal{L}^{\precsim}$*.*

Proof The lemma is proved by structural induction on $\alpha \in \mathcal{L}^{\precsim}$. We show that, for any $w \in W$ and $w' \in W'$, if $w\mathsf{E}w'$, then $\mathscr{W}, w \Vdash \alpha$ iff $\mathscr{W}', w' \Vdash \alpha$. For atomic propositions, and when $\alpha = \neg\beta$ or $\alpha = \beta_1 \vee \beta_2$, the proof is immediate. We consider the remaining two cases, namely when $\alpha = \Box_i\beta$ or $\alpha = \precsim_i\beta$.

Assume $\alpha = \Box_i\beta$ and let $\mathscr{W}, w \Vdash \Box_i\beta$. The proof is as for basic modal logic: Suppose $v' \in R'_i(w')$. Since $w\mathsf{E}w'$, there is some $v \in R_i(w)$ with $v\mathsf{E}v'$. Therefore $\mathscr{W}, v \Vdash \beta$, and hence $\mathscr{W}', v' \Vdash \beta$ by the induction hypothesis. It follows that $\mathscr{W}', w' \Vdash \Box_i\beta$. A symmetric argument applies if $\mathscr{W}', w' \Vdash \Box_i\beta$.

Assume $\alpha = \precsim_i\beta$ and let $\mathscr{W}, w \Vdash \precsim_i\beta$. Suppose $v' \in \min_{\prec} R'_i(w')$. Since $w\mathsf{E}w'$, there is some $v \in \min_{\prec} R_i(w)$ with $v\mathsf{E}v'$. Therefore $\mathscr{W}, v \Vdash \beta$, and hence $\mathscr{W}', v' \Vdash \beta$ by the induction hypothesis. It follows that $\mathscr{W}', w' \Vdash \precsim_i\beta$. A symmetric argument applies if $\mathscr{W}', w' \Vdash \precsim_i\beta$. ∎

We now turn to bisimulations between *R*-ordered models. As above, two worlds are bisimilar if they satisfy the same atomic information and their modal accessibility structures match, both in terms of accessible worlds and in terms of preference of accessibility.

Definition 10 **(*R*-ordered bisimulation)** Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ and $\mathscr{R}' = \langle W', R', V', \ll'\rangle$ be *R*-ordered models. An *R-ordered bisimulation* between $\mathscr{R}$ and $\mathscr{R}'$ is a non-empty binary relation $\mathsf{E} \subseteq W \times W'$ such that, whenever $w\mathsf{E}w'$, we have that:

1. For every $p \in \mathcal{P}$, $\mathscr{R}, w \Vdash p$ if and only if $\mathscr{R}', w' \Vdash p$;
2. if wR_iv, then there is v' in W' such that $v\mathsf{E}v'$ and $w'R'_iv'$, and

  - if $wv \in \min_{\ll_i} R_i^w$, then $w'v' \in \min_{\ll'_i} R_i'^{w'}$;
3. if $w'R'_iv'$, then there is v in W such that $v\mathsf{E}v'$ and wR_iv, and
  - if $w'v' \in \min_{\ll'_i} R_i'^{w'}$, then $wv \in \min_{\ll_i} R_i^w$.

Lemma 3 (**R-ordered bisimulation invariance lemma**) *If E is an R-ordered bisimulation between $\mathscr{R} = \langle W, R, V, \ll\rangle$ and $\mathscr{R}' = \langle W', R', V', \ll'\rangle$, $w \in W$ and $w' \in W'$, and $w\mathsf{E}w'$, then w and w' satisfy the same modal sentences in the extended language $\mathcal{L}^{\approxeq}$.*

Proof The proof is by structural induction on $\alpha \in \mathcal{L}^{\approxeq}$ and is similar to that of Lemma 2. We show that, for any $w, w' \in W$, if $w\mathsf{E}w'$, then $\mathscr{R}, w \Vdash \alpha$ iff $\mathscr{R}', w' \Vdash \alpha$. We only prove the case when $\alpha = \approxeq_i\beta$.

Assume $\alpha = \approxeq_i\beta$ and let $\mathscr{R}, w \Vdash \approxeq_i\beta$. Suppose $w'v' \in \min_{\ll'_i} R_i'^{w'}$. Since $w\mathsf{E}w'$, there is some $wv \in \min_{\ll_i} R_i^w$ with $v\mathsf{E}v'$. Therefore $\mathscr{R}, v \Vdash \beta$, and hence $\mathscr{R}', v' \Vdash \beta$ by the induction hypothesis. It follows that $\mathscr{R}', w' \Vdash \approxeq_i\beta$. A symmetric argument applies if $\mathscr{R}', w' \Vdash \approxeq_i\beta$. ∎

5 Expressive Power of $\approxeq$ Sentences

The relationship between $\mathcal{L}^{\approxeq}$ and $\mathcal{L}^{\approx}$, and between R-ordered and W-ordered models, can be made precise using our generalised bisimulations. We first show that $\mathcal{L}^{\approxeq}$ is at least as expressive as $\mathcal{L}^{\approx}$. That is, every class of models of some $\mathcal{L}^{\approx}$ formula is bisimilar to a class of models of some $\mathcal{L}^{\approxeq}$ formula [34]. Given a sentence $\alpha \in \mathcal{L}^{\approx}$, let $\alpha^{\approxeq}$ be the sentence obtained by replacing all occurrences of $\approx_i$ in α with $\approxeq_i$.

Definition 11 Let $\mathscr{W} = \langle W, R, V, \prec\rangle$. For any $u, v, w \in W$ such that wR_iu and wR_iv and $u \prec v$, let $wu \ll_i wv$. Then $\mathscr{R}_{\mathscr{W}} = \langle W, R, V, \ll\rangle$ is the *R-ordered model induced by $\mathscr{W}$*.

Lemma 4 *For any $\alpha \in \mathcal{L}^{\approx}$, $\mathscr{W} = \langle W, R, V, \prec\rangle$ and $w \in W$, $\mathscr{W}, w \Vdash \alpha$ if and only if in the R-ordered model $\mathscr{R}_{\mathscr{W}} = \langle W, R, V, \ll\rangle$ induced by $\mathscr{W}$, $\mathscr{R}_{\mathscr{W}}, w \Vdash \alpha^{\approxeq}$.*

Proof The proof is simple and proceeds by structural induction on the sentence α. ∎

Lemma 4 shows that, if α and β are not equivalent in $\mathcal{L}^{\approx}$ under the W-ordered semantics, then their translations $\alpha^{\approxeq}$ and $\beta^{\approxeq}$ are also not equivalent in $\mathcal{L}^{\approxeq}$ w.r.t. the R-ordered semantics. Further, if $(\mathscr{W}, w)$ and $(\mathscr{W}', w')$ are distinguishable by some $\alpha \in \mathcal{L}^{\approx}$, say, $\mathscr{W}, w \Vdash \alpha$ and $\mathscr{W}', w' \nVdash \alpha$, then $\mathscr{R}_{\mathscr{W}}$ and $\mathscr{R}'_{\mathscr{W}}$ are distinguishable by $\alpha^{\approxeq} \in \mathcal{L}^{\approxeq}$. Hence, $\mathcal{L}^{\approxeq}$ is at least as expressive as $\mathcal{L}^{\approx}$.

The converse of this result may not be as obvious to see, and translating R-ordered models to W-ordered models requires more care. The light switch example (Figure 1) shows that, even in the case of a single modality, there is no direct translation of a preference order on R to a preference order on W. There is no order on the two worlds w_1 and w_2 such that w_1 is the preferred result of toggling the light switch when the light is off, but w_2 is the preferred result when the light is on. A further problematic aspect is that R-ordered models allow for a preference order on each accessibility relation, whereas a W-ordered semantics assumes a single preference order on worlds.

Definition 12 Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ be an R-ordered model with single accessibility relation R_1. Let $W' = W \times W$; let $V'(uw) = V(w)$; let uvR'_1vw whenever vR_1w, and let $uv \prec u'v'$ whenever $uv \ll u'v'$. Then $\mathscr{W}_{\mathscr{R}} = \langle W', R', V', \prec\rangle$ is the *W-ordered model induced by $\mathscr{R}$*.

As an example, we apply Definition 12 to obtain the W-ordered models induced by the models of Figures 1 and 2, and depicted in Figures 7 and 8 respectively. Note that in Figure 7, $w_1w_2 \prec w_1w_1$ and $w_2w_1 \prec w_2w_2$, reflecting the intuition of normal execution of the action as an order on worlds. In Figure 8, the order on worlds is reversed, with $w_1w_1 \prec w_1w_2$ and $w_2w_2 \prec w_2w_1$, depicting the intuition of defeasible knowledge of the agent as an order on worlds.

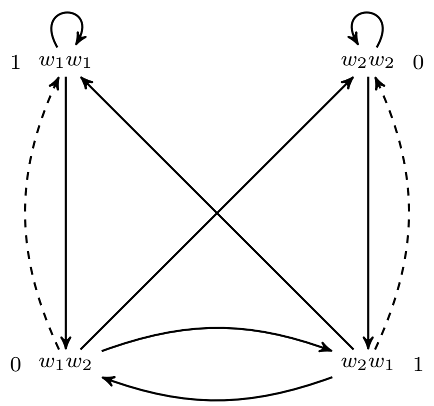

Fig. 7 The induced W-ordered model for one action (toggle) and one atom (on).

Theorem 2 *Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ be an R-ordered model with a single accessibility relation R_1 and let $\mathscr{W}_{\mathscr{R}} = \langle W, R, V, \prec\rangle$ be the W-ordered model induced by $\mathscr{R}$. Let $\mathscr{R}_{\mathscr{W}_{\mathscr{R}}} = \langle W', R', V', \ll'\rangle$ be the R-ordered model induced by $\mathscr{W}_{\mathscr{R}}$. Then there is a full bisimulation between $\mathscr{R}$ and $\mathscr{R}_{\mathscr{W}_{\mathscr{R}}}$, i.e., with domain W and range $W \times W$.*

Proof Let E be defined by: $w\mathsf{E}vw$ for all $v, w \in W$. We need to show that E is a full bisimulation relation. So, let $u, v \in W$. Then $v\mathsf{E}uv$.

1. It follows immediately from the construction of $\mathscr{R}_{\mathscr{W}_{\mathscr{R}}}$ that v and uv satisfy the same atomic propositions.

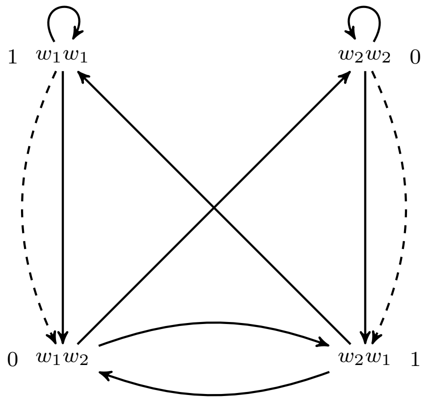

Fig. 8 The induced W-ordered model for one agent (A) and one atom (carbon).

2. Suppose vR_1w. It follows again from the construction of $\mathscr{R}_{\mathscr{W}_{\mathscr{R}}}$ that uvR'_1vw and $w\mathsf{E}vw$. Further, if $vw \in \min_{\ll_1} R_1^v$, then $vw \in \min_{\prec} R_1(uv)$, and hence $vw \in \min_{\ll'_1}(R'_1)^{uv}$.
3. Suppose uvR'_1vw. It again follows from the construction of $\mathscr{R}_{\mathscr{W}_{\mathscr{R}}}$ that vR_1w and $w\mathsf{E}vw$. Further, if $vw \in \min_{\ll'_1}(R'_1)^{uv}$, then $vw \in \min_{\prec} R_1(w)$, and hence $vw \in \min_{\ll_1} R_1^v$. ∎

We illustrate the construction of Theorem 2 by applying Definition 11 to the induced W-ordered model in Figure 7 to obtain the R-ordered model of Figure 9. In Figure 9, the dashed arrows represent the preference order $\ll'$. Theorem 2 then states that the R-ordered model of Figure 1 (with the order as described in the Introduction) is bisimilar to the R-ordered model of Figure 9. The construction is via the W-ordered model of Figure 7.

Similarly, the model of Figure 2 (again, with the order as described in the Introduction) is bisimilar to the R-ordered model of Figure 10, which is constructed via the W-ordered model of Figure 8.

Corollary 1 *$\mathcal{L}^{\approx}$ and $\mathcal{L}^{\approx}$ can distinguish between the same modal propositions when restricted to a single modality.*

Proof The bisimulation result of Theorem 2 shows that any R-ordered model is bisimilar to some R-ordered model induced by a W-ordered model. Lemma 3 ensures that bisimilar worlds satisfy the same modal sentences, and that bisimilar models can distinguish between the same modal properties. We need therefore consider only R-ordered models induced by some W-ordered model when reasoning about expressivity. The result then follows from Lemma 4. ∎

Corollary 1 may be seen as a negative result in the sense that, at least in the monomodal case, no richer language is obtained when substituting a preference order on the accessibility relation for the preference order on worlds. It is also

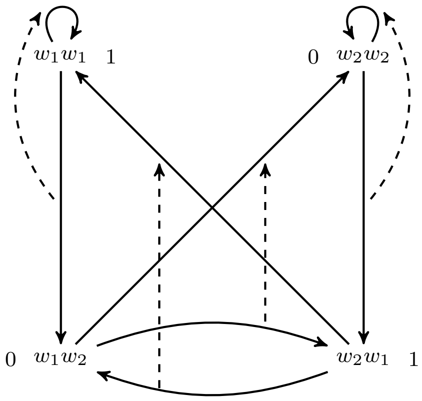

Fig. 9 The induced bisimilar R-ordered action model.

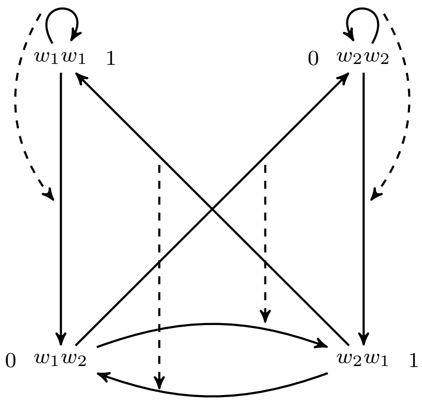

Fig. 10 The induced bisimilar R-ordered epistemic model.

clear that the results of Theorem 2 and Corollary 1 can be generalised to multi-modal languages if multiple preference relations on W are allowed.

What, then, has been gained? As we have argued, there are a number of contexts in which an order on the accessibility relation has an intuitive appeal. (A further example is the notion of defeasible roles in description-logic ontologies [18,21] which, too, is based on a preference on binary relations.) The induced W-ordered models of Definition 11 are technically useful, but intuitively hard to motivate. However, from an implementation perspective, we now know that a reasoner based on a W-ordered semantics such as the one we have defined for $\mathcal{L}^{\approx}$ [17] and its extension with conditionals [16], which run in polynomial space, suffices also for reasoning over R-ordered models. This establishes the following:

Corollary 2 *Satisfiability checking for monomodal $\mathcal{L}^{\approx}$ is decidable and is a* PSPACE-*complete problem.*

6 *R*-Based Conditionals

A framework for representing and reasoning with defeasibility would not be complete without an account of (defeasible) conditionals, which, historically, are at the heart of research in non-monotonic reasoning. In this section, we show how our semantic constructions are also fruitful in the definition of different types of conditional statements. Before doing so, we recall the KLM approach to conditionals of Kraus, Lehmann and Magidor [36,38], against which we shall check the proposals that we introduce below. Alternative approaches to conditionals in default reasoning have been investigated in the literature [5, 10,24,26,33], but here we shall content ourselves with a comparison to the KLM one, since it provides the gold standard by which non-monotonic implication is usually evaluated.

A *defeasible conditional* is a statement of the form $\alpha \leadsto \beta$, where α and β are sentences of the underlying logical language, of which the intuition is that "normally, if α, then β". We say that $\leadsto$ is a *preferential* conditional if it satisfies the following set of properties, alias postulates or Gentzen-style rules, as they are sometimes also referred to in the literature (below, $\models$ denotes validity in the subjacent logic):

$$(\text{Ref})\ \ \alpha \leadsto \alpha \qquad\qquad (\text{LLE})\ \ \frac{\models \alpha \leftrightarrow \beta,\ \alpha \leadsto \gamma}{\beta \leadsto \gamma}$$

$$(\text{And})\ \ \frac{\alpha \leadsto \beta,\ \alpha \leadsto \gamma}{\alpha \leadsto \beta \land \gamma} \qquad\qquad (\text{Or})\ \ \frac{\alpha \leadsto \gamma,\ \beta \leadsto \gamma}{\alpha \lor \beta \leadsto \gamma}$$

$$(\text{RW})\ \ \frac{\alpha \leadsto \beta,\ \models \beta \rightarrow \gamma}{\alpha \leadsto \gamma} \qquad\qquad (\text{CM})\ \ \frac{\alpha \leadsto \beta,\ \alpha \leadsto \gamma}{\alpha \land \beta \leadsto \gamma}$$

There is reasonable consensus in the literature that these properties correspond to the minimum requirements a conditional deemed as appropriate in a defeasible-reasoning setting ought to satisfy. They have been discussed at length [36–38] and we shall not do so here.

The intuition for the semantics of a sentence of the form $\alpha \leadsto \beta$ is that, in those most normal situations in which α holds, β also holds. In the propositional case, this is captured by placing a preference relation on the set of valuations [46] and evaluating $\alpha \leadsto \beta$ to true whenever the minimal α-valuations satisfy β. In what follows, we show how to lift this idea to more expressive languages in two different ways within our new semantics.

First, let *id* denote the *identity relation* on W and let us order its elements in the same way as for the other R-components. The intuition of doing so is that the most normal *id*-arrows correspond (implicitly) to the most normal worlds, i.e., we get an ordering on worlds induced by the ordering on the elements of the identity relation. With this, we can define our first candidate for a conditional in the following way. First, for every $\alpha \in \mathcal{L}^{\approxeq}$, let $id^{\alpha} := \{(w,w) \in id \mid \mathscr{R}, w \Vdash \alpha\}$. Then

- $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ if and only if $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\beta}$.

Proposition 2 *$\rightsquigarrow_{id}$ is a preferential conditional.*

Proof Let $\mathscr{R} = \langle W, R, V, \ll\rangle$ be an R-ordered model with $R = \langle R_1, \ldots, R_n, id\rangle$ and $\ll = \langle \ll_1, \ldots, \ll_n, \ll_{id}\rangle$, and let $\rightsquigarrow_{id}$ be defined as above.

(Ref): From $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\alpha}$, it follows that $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \alpha$.

(LLE): Let $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \gamma$ and $\models \alpha \leftrightarrow \beta$. Then $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\gamma}$. Moreover, since $\models \alpha \leftrightarrow \beta$, $id^{\alpha} = id^{\beta}$, and therefore $\min_{\ll_{id}} id^{\alpha} = \min_{\ll_{id}} id^{\beta}$. Hence $\mathscr{R} \Vdash \beta \rightsquigarrow_{id} \gamma$.

(And): Let $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ and $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \gamma$. Then $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\beta}$ and $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\gamma}$. Hence $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\beta} \cap id^{\gamma} = id^{\beta\wedge\gamma}$. Therefore $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta \wedge \gamma$.

(Or): Let $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \gamma$ and $\mathscr{R} \Vdash \beta \rightsquigarrow_{id} \gamma$. Then $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\gamma}$ and $\min_{\ll_{id}} id^{\beta} \subseteq id^{\gamma}$. Since $\min_{\ll_{id}} id^{\alpha} \cup \min_{\ll_{id}} id^{\beta} \supseteq \min_{\ll_{id}} id^{\alpha\vee\beta}$, we conclude that $\mathscr{R} \Vdash \alpha \vee \beta \rightsquigarrow_{id} \gamma$.

(RW): Let $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ and $\models \beta \rightarrow \gamma$. Then $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\beta}$ and $id^{\beta} \subseteq id^{\gamma}$. Hence $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \gamma$.

(CM): Let $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ and $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \gamma$. Then $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\beta}$ and $\min_{\ll_{id}} id^{\alpha} \subseteq id^{\gamma}$. From the latter, it follows that $\min_{\ll_{id}} id^{\alpha\wedge\gamma} \subseteq \min_{\ll_{id}} id^{\alpha}$. Hence $\mathscr{R} \Vdash \alpha \wedge \gamma \rightsquigarrow_{id} \beta$. ■

Notice that, semantically, $\rightsquigarrow_{id}$ relies on a single, absolute ordering. In some applications, notably in a multi-agent context, this is not particularly desirable, since it entails that objectivity and (an agent's) subjectivity collapse. As we shall see below, the versatility of R-ordered models shows us a way out.

In a multi-agent epistemic setting, one can enrich both $\mathcal{L}^{\approx}$ and $\mathcal{L}^{\approxeq}$ with modalities to express, for example, *conditional belief* of a given sentence α: $\approx_i^{\alpha}\beta$ or $\approxeq_i^{\alpha}\beta$. The intuition of such sentences is that, according to agent i's perspective, the most plausible amongst the α-worlds deemed possible are β-worlds, i.e., β is a *plausible consequence* of α, which corresponds to the intended meaning of a conditional statement of the form $\alpha \rightsquigarrow \beta$. In the case of $\mathcal{L}^{\approx}$, the semantics is as follows (below, $[\![\alpha]\!]^{\mathscr{W}}$ denotes the α-worlds in $\mathscr{W}$):

- $\mathscr{W}, w \Vdash \approx_i^{\alpha}\beta$ if and only if $\min_{\prec}(R_i(w) \cap [\![\alpha]\!]^{\mathscr{W}}) \subseteq [\![\beta]\!]^{\mathscr{W}}$

In the case of $\mathcal{L}^{\approxeq}$, let first, for each w and every α,

$$R_i^{w,\alpha} := \{(w,w') \mid (w,w') \in R_i \text{ and } \mathscr{R}, w' \Vdash \alpha\}$$

Then we define the semantics of $\approxeq_i^{\alpha}\beta$ as follows:

- $\mathscr{R}, w \Vdash \approxeq_i^{\alpha}\beta$ if and only if $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta}$

In either case, i.e., whether $\mathcal{L}^{\widetilde{\Box}}$ or $\mathcal{L}^{\overset{\approx}{\Box}}$ is assumed, we can define the conditional $\alpha \rightsquigarrow_i \beta$ as an abbreviation for the respective conditional defeasible modality, i.e., either $\alpha \rightsquigarrow_i \beta := \widetilde{\Box}_i^{\alpha}\beta$ or $\alpha \rightsquigarrow_i \beta := \overset{\approx}{\Box}_i^{\alpha}\beta$. Alternatively, we can add $\rightsquigarrow_i$ to the basic language and give it directly the semantics as defined in the above respective cases.

Proposition 3 *Each of* $\rightsquigarrow_i$ *is a preferential conditional.*

Proof Let $\mathscr{R} = \langle W, R, V, \ll\rangle$, where $R = \langle R_1, \ldots, R_n\rangle$ and $\ll = \langle \ll_1, \ldots, \ll_n\rangle$, and let $\rightsquigarrow_i$, $i = 1, \ldots, n$, be such that, for every $\alpha, \beta \in \mathcal{L}^{\overset{\approx}{\Box}}$ and every $w \in W$, $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \beta$ if and only if $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta}$.

(Ref): From $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\alpha}$ follows $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \alpha$.

(LLE): Let $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \gamma$ and $\models \alpha \leftrightarrow \beta$. Then $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\gamma}$. Moreover, since $\models \alpha \leftrightarrow \beta$, $R_i^{w,\alpha} = R_i^{w,\beta}$, and therefore $\min_{\ll_i} R_i^{w,\alpha} = \min_{\ll_i} R_i^{w,\beta}$. Hence $\mathscr{R}, w \Vdash \beta \rightsquigarrow_i \gamma$.

(And): Let $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \beta$ and $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \gamma$. Then $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta}$ and $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\gamma}$. Hence $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta} \cap R_i^{w,\gamma} = R_i^{w,\beta\wedge\gamma}$. Therefore $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \beta \wedge \gamma$.

(Or): Let $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \gamma$ and $\mathscr{R}, w \Vdash \beta \rightsquigarrow_i \gamma$. Then $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\gamma}$ and $\min_{\ll_i} R_i^{w,\beta} \subseteq R_i^{w,\gamma}$. Since $\min_{\ll_i} R_i^{w,\alpha} \cup \min_{\ll_i} R_i^{w,\beta} \supseteq \min_{\ll_i} R_i^{w,\alpha\vee\beta}$, we conclude that $\mathscr{R}, w \Vdash \alpha \vee \beta \rightsquigarrow_i \gamma$.

(RW): Let $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \beta$ and $\models \beta \rightarrow \gamma$. Then $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta}$ and $R_i^{w,\beta} \subseteq R_i^{w,\gamma}$. Hence $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \gamma$.

(CM): Let $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \beta$ and $\mathscr{R}, w \Vdash \alpha \rightsquigarrow_i \gamma$. Then $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\beta}$ and $\min_{\ll_i} R_i^{w,\alpha} \subseteq R_i^{w,\gamma}$. From the latter, it follows that $\min_{\ll_i} R_i^{w,\alpha\wedge\gamma} \subseteq \min_{\ll_i} R_i^{w,\alpha}$. Therefore $\mathscr{R}, w \Vdash \alpha \wedge \gamma \rightsquigarrow_i \beta$. ■

If, in addition to the KLM preferential properties, the defeasible conditional $\rightsquigarrow$ under consideration also satisfies the following Rational Monotonicity property [38], then it is said to be a *rational* conditional:

$$(\text{RM})\ \frac{\alpha \rightsquigarrow \beta, \quad \alpha \not\rightsquigarrow \neg\gamma}{\alpha \wedge \gamma \rightsquigarrow \beta}$$

There is a tacit agreement in the non-monotonic reasoning community that rational conditionals constitute a suitable class of implications for non-monotonic reasoning (both at the object and at the meta-level), one of the reasons stemming from its confluence with the AGM paradigm for belief revision [1,27,32]. It turns out we can ensure rationality of $\rightsquigarrow$ in a subclass of our R-ordered models, in which the relevant $\ll$-component is a *modular order*:

Definition 13 (Modular Order) Given a set X, $\ll_X \subseteq X \times X$ is modular if and only if there is a (strict) totally ordered set $\mathbb{Q}$, with the ordering denoted by $<$, and a ranking function $rk : X \longrightarrow \mathbb{Q}$ s.t. for every $x, y \in X$, $x \ll_X y$ if and only if $rk(x) < rk(y)$.

Proposition 4 *Let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be such that $R = \langle R_1, \ldots, R_n, id \rangle$ and $\ll = \langle \ll_1, \ldots, \ll_n, \ll_{id} \rangle$, and such that $\ll_{id}$ is a modular order. Then $\rightsquigarrow_{id}$ is a rational conditional.*

Proof Assume $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ but $\mathscr{R} \nVdash \alpha \rightsquigarrow_i \neg\gamma$. From the latter, it follows that there is $(w, w) \in \min_{\ll_{id}} id^{\alpha}$ such that $(w, w) \in id^{\gamma}$, i.e., $(w, w) \in id^{\alpha} \cap id^{\gamma} = id^{\alpha\wedge\gamma}$. Now let $(w', w') \in \min_{\ll_{id}} id^{\alpha\wedge\gamma}$. Since $(w, w) \in id^{\alpha\wedge\gamma}$, $(w, w) \not\ll_{id} (w', w')$ and then $rk(w', w') \leq rk(w, w)$. But since $(w', w') \in id^{\alpha}$, the case $rk(w', w') < rk(w, w)$ contradicts the fact that $(w, w) \in \min_{\ll_{id}} id^{\alpha}$. Hence $rk(w', w') = rk(w, w)$. This means $(w', w') \in \min_{\ll_{id}} id^{\alpha}$. From this and $\mathscr{R} \Vdash \alpha \rightsquigarrow_{id} \beta$ follows $(w', w') \in id^{\beta}$. Hence $\mathscr{R} \Vdash \alpha \wedge \gamma \rightsquigarrow_{id} \beta$. ■

The proof of the following result is along the lines of that for Proposition 4 and we leave it to the reader.

Proposition 5 *Let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be such that $R = \langle R_1, \ldots, R_n \rangle$ and $\ll = \langle \ll_1, \ldots, \ll_n \rangle$, and such that each $\ll_i$, $i = 1, \ldots, n$, is a modular order. Then each $\rightsquigarrow_i$ is a rational conditional.*

We shall leave an investigation of the notion of Rational Closure [38] in this more expressive setting for future work.

We close this section with a comment on the semantics of conditionals as discussed above. One of the limitations of standard preferential approaches [36, 46] and their multiple extensions [8, 9, 7, 10, 12, 14, 22, 23, 28–30] is the assumption of a single preference order on the set of worlds (or objects). This means there is no way to express the fact a given world w is more preferred than w' in a given context, while allowing for w' to be more preferred than w in a different context. With the above proposals for a semantics for $\rightsquigarrow$ within our framework, such a limitation carries over. Since one of the motivations behind our R-ordered models is precisely to allow for multiple preference relations, it would be natural to have multiple preferences on the set of worlds, too, each one giving rise to a context-based conditional. We have done a preliminary investigation on the matter for the case of context-based defeasible subsumption in description-logic ontologies [20], which can shed some light on how to tackle this issue in a modal setting.

7 Discussion and Related Work

We start by observing that the modal operators we have studied here do not aim at formalising the notion of *most*, as addressed in generalised quantifiers [43] and, more recently, in a modal context by Veloso et al. [48] and Askounis et al. [2]. Furthermore, our defeasible modalities are not about degrees of truth as has been studied in fuzzy logics, nor about degrees of possibility and necessity as addressed by possibilistic logics [25]. They rather relate to or generalise similar preferential construals we have studied previously [11, 18].

In a sense, the notions we investigated here can be seen as the qualitative counterpart of possibilistic modalities [41,42]. (We thank an anonymous referee for pointing this out to us.) There, each possible world w is associated with a *possibility distribution* $\pi_w : W \longrightarrow [0,1]$, the intuition of which is to capture the degree of likelihood (in terms of belief) of all possible worlds w.r.t. w. In that setting, the pairs (w, w') for which $\pi_w(w')$ is maximal correspond here to the most preferred pairs in a single accessibility relation. In this sense, there are strong links between *monomodal* $\approxeq$ and the preferential possibilistic semantics in the context of epistemic reasoning.

The observant reader would have noticed that the definition of R-ordered model (Definition 5) allows only for elements of the same accessibility relation R_i to be ordered (via the respective $\ll_i$). More generally, we could have defined $\ll$ as a relation on $\bigcup_{1\leq i\leq n} R_i \times \bigcup_{1\leq i\leq n} R_i$, so that we allow pairs (w, w') belonging to different R-components to be compared as well. An investigation of the philosophical and practical ramifications of this alternative definition is left for future work.

We have seen that one can obtain R-ordered models from W-ordered models by inducing an ordering on edges from the ordering on worlds. The result is an 'embedding' of $\backsim$ into $\mathcal{L}^{\approxeq}$. Conversely, in the monomodal case, we can obtain W-ordered models from R-ordered models by inducing an ordering on worlds from an ordering on edges. If we allow multiple preferences on worlds, the latter result can be generalised, thereby establishing that $\mathcal{L}^{\approxeq}$ and $\mathcal{L}^{\backsim}$ are equally expressive. This would have an interesting consequence, namely that the notions of 'normal effects' and 'normal executions' of actions are one and the same. This is justified by the dependence of the effects of an action (the worlds one 'lands' in) on the current state of the world (the 'departing' points). In other terms, talking about effects (tacitly) amounts to talking about pairs (w, w'), linking both a context of execution and the action's outcome. This feature just carries over when normality is considered.

In this work, we have not addressed the question as to what an appropriate notion of non-monotonic entailment for $\mathcal{L}^{\approxeq}$ is and have contented ourselves with the standard (Tarskian) definition, which is monotonic (and therefore not suitable in all contexts). The recent results by Booth et al. [7,8] extending the notion of Rational Closure [38] in a propositional setting may provide us with a springboard from which to investigate this matter in more expressive languages such as those we are interested in here.

8 Summary and Conclusion

The contributions of the present paper can be summarised as follows: (*i*) the motivation for and definition of a semantic structure allowing for the ordering of *pairs* of worlds (instead of worlds *tout court*, as is customary in traditional NMR formalisms and others); (*ii*) the definition of preferential filtration which comes in handy in showing the finite-model property of our extended modal logic (see Appendix A); (*iii*) a generalisation of bisimulation to the preferential case together with a result relating our new semantics to that we studied in previous work [17] and showing that, in the *monomodal* case, they are equivalent, and (*iv*) definitions of defeasible conditional statements on a language that is more expressive than *all* languages considered in recent extensions of the KLM approach [9, 14, 23, 29].

We have introduced a logic allowing for modal operators the intuition of which is to capture the qualitative idea of some *transitions* being more normal (or likely) than others. Our R-ordered models can be used to provide the extended language with an intuitive and elegant semantics. The resulting framework provides for an alternative formalisation of the notion of defeasible necessity we studied previously.

We have given examples, in an action, epistemic and deontic contexts, of what this semantic structure, as simple as it is, allows to represent (or give a meaning to) that one cannot straightforwardly achieve with standard Kripkean semantics.

Acknowledgments

We would like to thank the anonymous referees for their constructive comments and suggestions.

This work was partially funded by the National Research Foundation of South Africa (UIDs 81225 and 85482, IFR1202160021 and IFR2011032700018).

References

1. Alchourrón, C., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic **50**, 510–530 (1985)
2. Askounis, D., Koutras, C., Zikos, Y.: 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), no. 7519 in LNCS, pp. 41–53. Springer (2012)
3. Baltag, A., Smets, S.: Dynamic belief revision over multi-agent plausibility models. In: W. van der Hoek, M. Wooldridge (eds.) Proceedings of LOFT, pp. 11–24. University of Liverpool (2006)
4. Baltag, A., Smets, S.: 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), no. 3 in Texts in Logic and Games, pp. 13–60. Amsterdam University Press (2008)
5. Benferhat, S., Dubois, D., Prade, H.: Possibilistic and standard probabilistic semantics of conditional knowledge bases. Journal of Logic and Computation **9**(6), 873–895 (1999)
6. Blackburn, P., Benthem, J., Wolter, F.: Handbook of Modal Logic. Elsevier North-Holland (2006)
7. Booth, R., Casini, G., Meyer, T., Varzinczak, I.: On the entailment problem for a logic of typicality. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI) (2015)

8. Booth, R., Meyer, T., Varzinczak, I.: 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), no. 7519 in LNCS, pp. 107–119. Springer (2012)
9. Booth, R., Meyer, T., Varzinczak, I.: A propositional typicality logic for extending rational consequence. In: E. Fermé, D. Gabbay, G. Simari (eds.) Trends in Belief Revision and Argumentation Dynamics, *Studies in Logic – Logic and Cognitive Systems*, vol. 48, pp. 123–154. King's College Publications (2013)
10. Boutilier, C.: Conditional logics of normality: A modal approach. Artificial Intelligence **68**(1), 87–154 (1994)
11. Britz, K., Casini, G., Meyer, T., Varzinczak, I.: Preferential role restrictions. In: Proceedings of the 26th International Workshop on Description Logics, pp. 93–106 (2013)
12. Britz, K., Heidema, J., Meyer, T.: Semantic preferential subsumption. In: J. Lang, G. Brewka (eds.) Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR), pp. 476–484. AAAI Press/MIT Press (2008)
13. Britz, K., Meyer, T., Varzinczak, I.: Preferential reasoning for modal logics. Electronic Notes in Theoretical Computer Science **278**, 55–69 (2011). Proceedings of the 7th Workshop on Methods for Modalities (M4M'2011)
14. Britz, K., Meyer, T., Varzinczak, I.: Semantic foundation for preferential description logics. In: D. Wang, M. Reynolds (eds.) Proceedings of the 24th Australasian Joint Conference on Artificial Intelligence, no. 7106 in LNAI, pp. 491–500. Springer (2011)
15. Britz, K., Meyer, T., Varzinczak, I.: Normal modal preferential consequence. In: M. Thielscher, D. Zhang (eds.) Proceedings of the 25th Australasian Joint Conference on Artificial Intelligence, no. 7691 in LNAI, pp. 505–516. Springer (2012)
16. Britz, K., Varzinczak, I.: From KLM-style conditionals to defeasible modalities, and back. Journal of Applied Non-Classical Logics To appear
17. Britz, K., Varzinczak, I.: Defeasible modalities. In: Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK), pp. 49–60 (2013)
18. Britz, K., Varzinczak, I.: Introducing role defeasibility in description logics. In: L. Michael, A. Kakas (eds.) Proceedings of the 15th European Conference on Logics in Artificial Intelligence (JELIA), no. 10021 in LNCS, pp. 174–189. Springer (2016)
19. Britz, K., Varzinczak, I.: Preferential modalities revisited. In: Proceedings of the 16th International Workshop on Nonmonotonic Reasoning (NMR) (2016)
20. Britz, K., Varzinczak, I.: Context-based defeasible subsumption for $d\mathcal{SROIQ}$. In: Proceedings of the 13th International Symposium on Logical Formalizations of Commonsense Reasoning (2017)
21. Britz, K., Varzinczak, I.: Towards defeasible $d\mathcal{SROIQ}$. In: Proceedings of the 30th International Workshop on Description Logics, vol. 1879. CEUR Workshop Proceedings (2017)
22. Casini, G., Meyer, T., Moodley, K., Sattler, U., Varzinczak, I.: Introducing defeasibility into OWL ontologies. In: M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, S. Staab (eds.) Proceedings of the 14th International Semantic Web Conference (ISWC), no. 9367 in LNCS, pp. 409–426. Springer (2015)
23. Casini, G., Straccia, U.: Rational closure for defeasible description logics. In: T. Janhunen, I. Niemelä (eds.) Proceedings of the 12th European Conference on Logics in Artificial Intelligence (JELIA), no. 6341 in LNCS, pp. 77–90. Springer-Verlag (2010)
24. Crocco, G., Lamarre, P.: 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), pp. 565–571. Morgan Kaufmann Publishers (1992)
25. Dubois, D., Lang, J., Prade, H.: Possibilistic logic. In: D. Gabbay, C. Hogger, J. Robinson (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3, pp. 439–513. Oxford University Press (1994)
26. Friedman, N., Halpern, J.: Plausibility measures and default reasoning. Journal of the ACM **48**(4), 648–685 (2001)
27. Gärdenfors, P., Makinson, D.: Nonmonotonic inference based on expectations. Artificial Intelligence **65**(2), 197–245 (1994)

28. Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.: Preferential description logics. In: N. Dershowitz, A. Voronkov (eds.) Logic for Programming, Artificial Intelligence, and Reasoning (LPAR), no. 4790 in LNAI, pp. 257–272. Springer (2007)
29. Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.: A non-monotonic description logic for reasoning about typicality. Artificial Intelligence **195**, 165–202 (2013)
30. Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.: Semantic characterization of rational closure: From propositional logic to description logics. Artificial Intelligence **226**, 1–33 (2015)
31. Hansson, B.: An analysis of some deontic logics. Noûs **3**, 373–398 (1969)
32. Hansson, S.: A Textbook of Belief Dynamics: Theory Change and Database Updating. Kluwer Academic Publishers (1999)
33. Hawthorne, J.: Nonmonotonic conditionals that behave like conditional probabilities above a threshold. Journal of Applied Logic **5**(4), 625–637 (2007)
34. Hodges, W.: Model Theory. Cambridge University Press (1993)
35. Katsuno, H., Mendelzon, A.: Propositional knowledge base revision and minimal change. Artificial Intelligence **3**(52), 263–294 (1991)
36. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**, 167–207 (1990)
37. Lehmann, D.: Another perspective on default reasoning. Annals of Mathematics and Artificial Intelligence **15**(1), 61–82 (1995)
38. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence **55**, 1–60 (1992)
39. Lewis, D.: Counterfactuals. Blackwell (1973)
40. Lewis, D.: Semantic analyses for dyadic deontic logic. In: S. Stenlund (ed.) Logical Theory and Semantic Analysis, pp. 1–14. D. Reidel Publishing Company (1974)
41. Liau, C.J.: On the possibility theory-based semantics for logics of preference. International Journal of Approximate Reasoning **20**(2), 173–190 (1999)
42. Liau, C.J., Lin, B.P.: Possibilistic reasoning–a mini-survey and uniform semantics. Artificial Intelligence **88**(1-2), 163–193 (1996)
43. Lindström, P.: First-order predicate logic with generalized quantifiers. Theoria **32**, 286–195 (1966)
44. Makinson, D.: Five faces of minimality. Studia Logica **52**, 339–379 (1993)
45. Makinson, D.: Bridges from Classical to Nonmonotonic Logic, *Texts in Computing*, vol. 5. King's College Publications (2005)
46. Shoham, Y.: Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence. MIT Press (1988)
47. Stalnaker, R.: A theory of conditionals. In: N. Rescher (ed.) Studies in Logical Theory, pp. 98–112. Blackwell (1968)
48. Veloso, P., Veloso, S., Viana, J., de Freitas, R., Benevides, M., Delgado, C.: On vague notions and modalities: a modular approach. Logic Journal of the IGPL **18**(3), 381–402 (2009)

A Proof of Theorem 1

In the following, with $\Diamond\!\!\!\approx_i$ we denote the dual of $\approx_i$ in the usual sense, i.e., $\Diamond\!\!\!\approx_i \alpha := \neg \approx_i \neg\alpha$.

Definition 14 (Subsentence) Let $\alpha \in \mathcal{L}^{\approx}$. The set of *subsentences* of α, denoted by $sub(\alpha)$, is defined inductively by:

- If $\alpha = \top$ or $\alpha = \bot$, then $sub(\alpha) := \{\alpha\}$;
- If $\alpha = p \in \mathcal{P}$, then $sub(\alpha) := \{p\}$;
- If $\alpha = \beta \wedge \gamma$ or $\alpha = \beta \vee \gamma$ or $\alpha = \beta \to \gamma$, then $sub(\alpha) := \{\alpha\} \cup sub(\beta) \cup sub(\gamma)$;
- If $\alpha = \neg\beta$ or $\alpha = \Diamond_i\beta$ or $\alpha = \Box_i\beta$ or $\alpha = \Diamond\!\!\!\approx_i\beta$ or $\alpha = \approx_i\beta$, then $sub(\alpha) := \{\alpha\} \cup sub(\beta)$.

Definition 15 (Closure under Subsentences) Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$. We say $\mathcal{C}$ is *closed under its subsentences* if and only if $\bigcup\{sub(\alpha) \mid \alpha \in \mathcal{C}\} \subseteq \mathcal{C}$.

Definition 16 ($\mathcal{C}$-type) Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$ and let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be an R-ordered model. For every $w \in W$,

$$t_{\mathcal{C}}(w) := \{\alpha \in \mathcal{C} \mid \mathscr{R}, w \Vdash \alpha\}$$

is the $\mathcal{C}$-type of w in $\mathscr{R}$.

Definition 17 ($\mathcal{C}$-equivalence) Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$ be finite and let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be an R-ordered model. For every $w, w' \in W$, $w \simeq_{\mathcal{C}} w'$ if and only if $t_{\mathcal{C}}(w) = t_{\mathcal{C}}(w')$. (It is easy to see that $\simeq_{\mathcal{C}}$ is an equivalence relation on W.) With $[w]_{\mathcal{C}} := \{w' \mid w \simeq_{\mathcal{C}} w'\}$, we denote the $\mathcal{C}$-equivalence class of $w \in W$.

Definition 18 ($\mathcal{C}$-filtration) Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$ be finite and let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be an R-ordered model. Moreover, let:

- $W' := W/\simeq_{\mathcal{C}}$ (i.e., $W' = \{[w]_{\mathcal{C}} \mid w \in W\}$);
- $R' := \langle R'_1, \ldots, R'_n \rangle$, where $R'_i := \{([w]_{\mathcal{C}}, [w']_{\mathcal{C}}) \mid (w, w') \in R_i\}$, $i = 1, \ldots, n$;
- $V' : W' \longrightarrow \{0,1\}^{\mathcal{P}}$, with $V'([w]_{\mathcal{C}}) = V(w)$;
- $\ll' = \langle \ll'_1, \ldots, \ll'_n \rangle$, where $\ll'_i := \{([w]_{\mathcal{C}}[v]_{\mathcal{C}}, [w']_{\mathcal{C}}[v']_{\mathcal{C}}) \mid (wv, w'v') \in \ll_i$ and either $w \notin [w']_{\mathcal{C}}$ or $v \notin [v']_{\mathcal{C}}\}$, $i = 1, \ldots, n$. (It is not hard to check that each $\ll'_i$ is a strict well-founded partial order.)

With $\mathscr{R}' := \langle W', R', V', \ll' \rangle$ we denote the $\mathcal{C}$-filtration of $\mathscr{R}$.

In Definition 18, the purpose of the clause "either $w \notin [w']_{\mathcal{C}}$ or $v \notin [v']_{\mathcal{C}}$" is to prevent two indistinguishable (w.r.t. $\mathcal{C}$) pairs of worlds from inducing a reflexive $\ll'_i$-edge, which would violate the stated condition that $\ll'_i$ must be a strict partial order.

Lemma 5 (Minimality Preservation) *Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$ be finite and closed under subsentences, let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be an R-ordered model, and let $\mathscr{R}' = \langle W', R', V', \ll' \rangle$ be the $\mathcal{C}$-filtration of $\mathscr{R}$. Then, for $i = 1, \ldots, n$, (i) if $(w, v) \in \min_{\ll_i} R_i$, then $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \in \min_{\ll'_i} R'_i$, and (ii) if $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \in \min_{\ll'_i} R'_i$, then there are $w' \in [w]_{\mathcal{C}}$ and $v' \in [v]_{\mathcal{C}}$ such that $(w', v') \in \min_{\ll_i} R_i$.*

Proof
Showing (i): Assume $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \notin \min_{\ll'_i} R'_i$. If $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \notin R'_i$, then $(w, v) \notin R_i$ and we are done. If there is $([w']_{\mathcal{C}}, [v']_{\mathcal{C}}) \in R'_i$ such that $([w']_{\mathcal{C}}, [v']_{\mathcal{C}}) \ll'_i ([w]_{\mathcal{C}}, [v]_{\mathcal{C}})$, then, by the construction of $\ll'_i$, we have $(x, y) \ll_i (w, v)$ for some $x \in [w]_{\mathcal{C}}$ and $y \in [v]_{\mathcal{C}}$. Hence $(w, v) \notin \min_{\ll_i} R_i$.

Showing (ii): Let $X := \{(w', v') \mid w' \in [w]_{\mathcal{C}}$ and $v' \in [v]_{\mathcal{C}}\}$. Assume $X \cap \min_{\ll_i} R_i = \emptyset$. Then for every $(w', v') \in X$ there must be $(x, y) \in R_i$ such that $(x, y) \ll_i (w', v')$. Moreover, $x \notin [w]_{\mathcal{C}}$ or $y \notin [v]_{\mathcal{C}}$, otherwise $(x, y) \in X$. Hence $([x]_{\mathcal{C}}, [y]_{\mathcal{C}}) \ll'_i ([w']_{\mathcal{C}}, [v']_{\mathcal{C}})$, by the construction of $\mathscr{R}'$. Since $[w']_{\mathcal{C}} = [w]_{\mathcal{C}}$ and $[v']_{\mathcal{C}} = [v]_{\mathcal{C}}$, it follows that $([x]_{\mathcal{C}}, [y]_{\mathcal{C}}) \ll'_i ([w]_{\mathcal{C}}, [v]_{\mathcal{C}})$, and therefore $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \notin \min_{\ll'_i} R'_i$. ■

Lemma 6 (Satisfaction Preservation) *Let $\mathcal{C} \subseteq \mathcal{L}^{\approx}$ be finite and closed under subsentences, let $\mathscr{R} = \langle W, R, V, \ll \rangle$ be an R-ordered model, and let $\mathscr{R}' = \langle W', R', V', \ll' \rangle$ be the $\mathcal{C}$-filtration of $\mathscr{R}$. For every $w \in W$ and every $\alpha \in \mathcal{C}$, $\mathscr{R}, w \Vdash \alpha$ if and only if $\mathscr{R}', [w]_{\mathcal{C}} \Vdash \alpha$.*

Proof The proof is by induction on the structure of the sentence α. Below we only show the case of $\approx$ via its dual $\diamondsuit$. (The classical cases are as usual.)

Assume that $\alpha = \diamondsuit_i \beta$. Since $\mathcal{C}$ is closed, we have $\beta \in \mathcal{C}$, and thus the induction hypothesis applies to β.

For the only-if part, let $\mathscr{R}, w \Vdash \diamondsuit_i \beta$. Then there is $w' \in W$ such that $(w, w') \in R_i$, $(w, w') \in \min_{\ll_i} R_i^w$ and $\mathscr{R}, w' \Vdash \beta$. We have $([w]_{\mathcal{C}}, [w']_{\mathcal{C}}) \in R'_i$, since $w \in [w]_{\mathcal{C}}$ and $w' \in [w']_{\mathcal{C}}$. By induction, we get $\mathscr{R}', [w']_{\mathcal{C}} \Vdash \beta$. By Lemma 5, $([w]_{\mathcal{C}}, [w']_{\mathcal{C}}) \in \min_{\ll'_i} R_i^{[w]_{\mathcal{C}}}$. Hence $\mathscr{R}', [w]_{\mathcal{C}} \Vdash \diamondsuit_i \beta$.

For the if part, let $\mathscr{R}', [w]_{\mathcal{C}} \Vdash \diamondsuit_i \beta$. Then there is $[v]_{\mathcal{C}} \in W'$ such that $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \in R'_i$, $([w]_{\mathcal{C}}, [v]_{\mathcal{C}}) \in \min_{\ll'_i} R_i'^{[w]_{\mathcal{C}}}$ and $\mathscr{R}', [v]_{\mathcal{C}} \Vdash \beta$. By induction, we get $\mathscr{R}, v \Vdash \beta$. Moreover, by Lemma 5, there are $w' \in [w]_{\mathcal{C}}$ and $v' \in [v]_{\mathcal{C}}$ such that $(w', v') \in \min_{\ll_i} R_i^{w'}$, from which follows $(w', v') \in R_i$. Since $v \simeq_{\mathcal{C}} v'$ and $\beta \in \mathcal{C}$, $\mathscr{R}, v \Vdash \beta$ implies $\mathscr{R}, v' \Vdash \beta$. Hence we have $\mathscr{R}, w' \Vdash \diamondsuit_i \beta$. Since $w \simeq_{\mathcal{C}} w'$ and $\diamondsuit_i \beta \in \mathcal{C}$, we get $\mathscr{R}, w \Vdash \diamondsuit_i \beta$. ■

Lemma 7 *If $\alpha \in \mathcal{L}^{\approx}$ is satisfiable, then it has a finite model.*

Proof Let $\mathscr{R}$ be such that $\mathscr{R} \Vdash \alpha$, and let $\mathcal{C} = sub(\alpha)$. From finiteness of α, it follows that $\mathcal{C}$ is finite, too. Hence Definition 18 guarantees the existence of a $\mathcal{C}$-filtration $\mathscr{R}'$ of $\mathscr{R}$ having a finite set of worlds. It remains to show that $\mathscr{R}'$ is a model of α. Let $w \in W$ be such that $\mathscr{R}, w \Vdash \alpha$. Since $\alpha \in \mathcal{C}$, by Lemma 6 it follows that $\mathscr{R}', [w]_{\mathcal{C}} \Vdash \alpha$. ■

The proof of Theorem 1 follows immediately from Lemma 7.