A modal perspective on defeasible reasoning

K. Britz[1], J. Heidema and W.A. Labuschagne

abstract. We introduce various new supraclassical entailment relations for defeasible reasoning and investigate some of their relationships. The contrapositive of the usual nonmonotonic preferential entailment turns out to be a monotonic supraclassical entailment. Analogously to these two, a nonmonotonic and its contrapositive monotonic supraclassical entailment can be defined by using modal operators relative to a reflexive accessibility relation instead of a preference order. Then again, by using the strict variant of a preference order on worlds as accessibility relation, modal versions of the usual preferential entailment and its contrapositive are obtained. Proof-theoretic aspects of modal preferential entailment is discussed briefly, leading to a characterization of modular Gödel-Löb logics. The link established between preferential and modal logics make existing modal proof and satisfiability algorithms available for use in preferential logics.

Keywords: modal logic; preferential logic; nonmonotonic logic; supraclassical entailment; defeasible reasoning.

1 Introduction

The heart of logic is entailment – a relation between information-bearers X and Y, induced by a relation E between representations $P(X)$ and $Q(Y)$ of X and Y. E represents the idea that X lends credence (maybe even truth) to Y. Classical entailment comes in at least two forms: syntactic and semantic. In syntactic entailment $X = P(X)$ and $Y = Q(Y)$ are sentences, while $E = \vdash$, formal deduction according to a set of syntactic rules, and we write $X \vdash Y$. When sentence X semantically entails sentence Y, written $X \models Y$, $P(X) = \mathbf{X} = \mathrm{Mod}(X)$, the set of models of X, $Q(Y) = \mathbf{Y} = \mathrm{Mod}(Y)$, and $E = \subseteq$, set-theoretical inclusion. Classical entailment, in either of its two forms, preserves truth: If X is true in an interpretation, then Y is also true in that interpretation.

In *supraclassical* entailment we allow more pairs (X, Y) into the entailment relation than can be justified classically. This is done by suitable choices, determined by meta-information beyond that carried by X and Y, of $P(X)$, $Q(Y)$ and E, and then stipulating "X entails Y" to mean $P(X)EQ(Y)$. Then the truth of X induces the plausibility, credibility, or likelihood, but not always the truth, of Y. The entailment is therefore *defeasible*; it may have a counterexample, an (unexpected) interpretation under

[1]Corresponding author

Paper submitted to Advances in Modal Logic 2006

which X is true, but Y is false. One of the well-studied defeasible ways of reasoning is known as nonmonotonic logic with semantics co-determined by a preference order on worlds or interpretations – *preferential logic*, for short [13, 9].

In Section 2 we introduce the contrapositive $\mid\!\sim^*$ of the usual preferential entailment $\mid\!\sim$. It turns out to have merits dual to those of $\mid\!\sim$, to be monotonic, and to be apt for abduction. Analogously to the dual pair $\mid\!\sim$ and $\mid\!\sim^*$, in Section 3 we discuss two modal defeasible entailments based on $\Box$ and $\Diamond$ with reflexive accessibility, again nonmonotonic and monotonic respectively. Section 4 demonstrates (now differently from the modal approach of Section 3) that $\mid\!\sim$ and $\mid\!\sim^*$ have exact modal equivalents. In Section 5 the modal defeasible entailment relations of Section 4 are positioned in a broader context of modal logics, related to Gödel-Löb logics, and shown to open up the well-developed decision procedures for modal and description logics for use in preferential logics.

2 Preferential logics

In a preferential logic one assumes as given an order relation on a set of possible worlds $\mathbf{W}$, for example a total preorder (a connected, reflexive, transitive, binary relation) or a partial order on the set of all worlds or all interpretations of the language. In the analysis of *rational* preferential reasoning, the partial order is assumed to be *modular*, i.e. for all u, v, w in $\mathbf{W}$, if u and v are incomparable and u is strictly below w, then v is also strictly below w. We also assume that the order relation is transitive and *Noetherian* (and hence smooth), i.e. there is no infinite strictly ascending chain of worlds. In the presence of transitivity, the Noetherian property is equivalent to the following condition: For every nonempty subset $\mathbf{U}$ of $\mathbf{W}$ and $u \in \mathbf{U}$ there is an element $v \in \mathbf{U}$, maximal in $\mathbf{U}$, with v greater than or equal to u.

Although this is not in general the case [9], we shall identify possible worlds with interpretations of the language. The intuitive idea captured by the preference order is then that interpretations higher up (greater) in the order are more preferred, more normal, more likely to occur in the context under consideration, than those lower down. For historical reasons the order is often inverted, i.e. "lower down/smaller = better" (see, for example, [1]), but we will not follow that convention in this paper.

In preferential logic with preference order $\leq$ on $\mathbf{W}$ the classical entailment relation $X \models Y$, i.e. $\mathbf{X} \subseteq \mathbf{Y}$, is expanded to a larger set of pairs (X, Y) by shrinking $\mathbf{X}$ to a smaller set, the set $P(X)$ of most preferred models of X, fitting into more different sets $\mathbf{Y}$. We define the defeasible entailment relation $\mid\!\sim$ by

$$X \mid\!\sim Y \text{ iff } P(X) \subseteq \mathbf{Y},$$

where

$$\begin{aligned} P(X) &:= \text{MaxMod}(X) \\ &:= \{w \in \mathbf{X} \mid \neg(\exists w' \in \mathbf{X})(w \leq w' \wedge w' \not\leq w)\}. \end{aligned}$$

MaxMod(X) is the set of maximal models of X – those models w of X for which there is *no* model of X strictly higher up in the preference order $\leq$ than w, i.e. strictly preferred to w [13, p.76]. So, a counterexample defeating $X \mid\!\sim Y$, a model of X which is not a model of Y, cannot be among the most preferred models of X, and must be rather "abnormal" – for the cognoscenti: like a world with Tweety unable to fly.

For a fixed premise X, the set $\{Y \mid X \mid\!\sim Y\}$ of consequences under $\mid\!\sim$ has good properties: it is a filter in the Lindenbaum-Tarski algebra of the language, i.e. is closed under classical conjunction $\wedge$ and classical entailment $\models$. (It is a principal filter if $P(X)$ is the set of models of a single sentence, as always happens for a finitely generated propositional language.) But, for a fixed consequence Y, the set of its premises $\{X \mid X \mid\!\sim Y\}$ under $\mid\!\sim$ merits no acclamation: it is not an ideal, i.a. because $\mid\!\sim$ is *nonmonotonic*; $X \mid\!\sim Y$ does not always ensure that $X \wedge X' \mid\!\sim Y$, so the set of premises of Y under $\mid\!\sim$ is not closed downward in the Lindenbaum-Tarski algebra. Could there be some variant, say $\mid\!\sim^*$, of $\mid\!\sim$ for which the premises $\{X \mid X \mid\!\sim^* Y\}$ of a fixed Y *do* always form an ideal – maybe at the price of not always having $\{Y \mid X \mid\!\sim^* Y\}$ a filter? The answer is yes – taking $\mid\!\sim^*$ to be the contrapositive of $\mid\!\sim$ performs the feat.

Classically, $X \models Y$ and its contrapositive $\neg Y \models \neg X$ are equivalent. This is not true for $\mid\!\sim$. Let us investigate the contrapositive of $X \mid\!\sim Y$ (where $\mathbf{W}$ is the set of all interpretations):

$$\begin{aligned} \neg Y \mid\!\sim \neg X \quad & \text{iff} \quad P(\neg Y) \subseteq \mathbf{W} - \mathbf{X} \\ & \text{iff} \quad \mathbf{X} \subseteq \mathbf{W} - P(\neg Y) \\ & \text{iff} \quad \mathbf{X} \subseteq \mathbf{Y} \cup [(\mathbf{W} - \mathbf{Y}) - P(\neg Y)]. \end{aligned}$$

If we now define

$$Q(Y) := \mathbf{Y} \cup [(\mathbf{W} - \mathbf{Y}) - P(\neg Y)],$$

adding to $\mathbf{Y}$ those models of $\neg Y$ which are not most preferred, then we can naturally define $\mid\!\sim^*$ by

$$X \mid\!\sim^* Y \text{ iff } \mathbf{X} \subseteq Q(Y).$$

Note that whereas $\mid\!\sim$ expands the relation $X \models Y$ by shrinking $\mathbf{X}$ to $P(X)$, $\mid\!\sim^*$ does this (differently) by dilating $\mathbf{Y}$ to $Q(Y)$. The intuition behind this is that, should X (against expectations) have a model not in $\mathbf{Y}$ (i.e. a counterexample to $X \models Y$), then this model (counterexample) is "abnormal", being not amongst the most preferred models of $\neg Y$. The consequences $\{Y \mid X \mid\!\sim^* Y\}$ of a fixed X under $\mid\!\sim^*$ do not constitute a filter in the Lindenbaum-Tarski algebra, but now the premises $\{X \mid X \mid\!\sim^* Y\}$ of a fixed Y *do* form an ideal, and $\mid\!\sim^*$ is *monotonic*. So, in this sense, $\mid\!\sim$ and its contrapositive $\mid\!\sim^*$ seem to have equal claims to being among the useful and acceptable defeasible expansions of $\models$, maybe apt in different contexts.

According to C.S. Peirce, non-deductive reasoning is either inductive or abductive [5]. His classification proceeds in terms of a model of scientific inquiry, leading to the characterization of abduction as hypothesis generation and induction as hypothesis evaluation by testing against reality through selected predictions. Peirce's model, while it has the virtue of simplicity, does not compel wide acceptance, and in particular the notion that induction is a process of testing hypotheses departs markedly from the Carnapian notion of inductive generalization from particular facts. Given that inductive generalizations (such as "All birds can fly") are usually defeasible, i.e. admit exceptions, it would seem that they should be formalized not via universal quantification but via a defeasible entailment relation. We would suggest, therefore, that the filter of consequences of X under $\mid\!\sim$ is the appropriate representation of the inductive generalization based on X. Reversing figure and ground, we suggest that the ideal of premises of Y under $\mid\!\sim^*$ is a mathematically precise formalization of the process of forming explanatory hypotheses that Peirce viewed as a rather mysterious guessing game or act of insight, namely abduction to an explanation of observation Y.

3 Defeasible reasoning using modal operators

In the previous section we explained how the classical entailment relation $\models$ can be expanded in two different ways to defeasible, but plausible, entailments $\mid\!\sim$ and $\mid\!\sim^*$, using a given preference order and induced shrinking and dilating operators P and Q on sets of models. In this section we investigate whether something similar can be done, but now letting P and Q be (induced by) modal operators $\Box$ and $\Diamond$ as yielded by a given accessibility relation on $\mathbf{W}$, instead of a preference order. The accessibility relation R will (only in this section) be assumed to be at least reflexive.

We note that $\Box$ shrinks model sets from $\mathbf{X}$ to $P(X) = \mathrm{Mod}(\Box X)$:

$$\begin{aligned}\mathrm{Mod}(\Box X) &= \{w \in \mathbf{W} \mid (\forall w')(wRw' \to w' \in \mathbf{X})\}\\ &= \{w \mid R[\{w\}] \subseteq \mathbf{X}\},\end{aligned}$$

a subset of $\mathbf{X}$, since R is reflexive and $w \in R[\{w\}]$. This is the subset consisting of those elements of $\mathbf{X}$ which are R-related to no world outside $\mathbf{X}$. Just as in the previous section, where $P(X)$ represented a logically stronger statement than X, here $\Box X$ ("necessarily, X") is logically stronger (has fewer models) than X. The consequences defeasibly and nonmonotonically entailed by fixed premise X, $\{Y \mid \Box X \models Y\}$, form a principal filter in the Lindenbaum-Tarski algebra. We note that if R is also transitive, then $\Box X \models Y$ is equivalent to $\Box X \models \Box Y$.

One possible reading of $\Box$ is as an *epistemic operator*. On this reading, the role of $\Box$ is similar to that of the epistemic operator K in description logics, where it is used to formalize the semantics of procedural rules in knowledge systems [4].

The modal operator $\Diamond$ dilates model sets from $\mathbf{Y}$ to $Q(Y) = \mathrm{Mod}(\Diamond Y)$:

$$\begin{aligned}\mathrm{Mod}(\Diamond Y) &= \{w \in \mathbf{W} \mid (\exists w')(wRw' \wedge w' \in \mathbf{Y})\} \\ &= R^{-1}[\mathbf{Y}],\end{aligned}$$

a superset of $\mathbf{Y}$, since R is reflexive. This is the set of all those worlds which are R-related to some world in $\mathbf{Y}$. $Q(Y)$ represents a logically weaker statement $\Diamond Y$ ("possibly, Y") than Y. The premises defeasibly, but monotonically, entailing a fixed consequence Y, $\{X \mid X \models \Diamond Y\}$, form a principal ideal in the Lindenbaum-Tarski algebra. Again, we note that if R is also transitive, then $X \models \Diamond Y$ is equivalent to $\Diamond X \models \Diamond Y$.

So, the shrinking and dilating operators P and Q on model sets, induced by the modal operators $\Box$ and $\Diamond$ on propositions, behave similarly to the P and Q of the previous section, there induced by a preference order. The two corresponding modal defeasible entailment relations, nonmonotonic $\Box X \models Y$ and monotonic $X \models \Diamond Y$, are not equivalent – they are not classical contrapositives, in spite of the duality $\Diamond = \neg\Box\neg$!

4 Modal preferential entailment

Preferential entailment is induced in a very specific way by a preference relation on $\mathbf{W}$. In contrast, an accessibility relation for a modal logic can, in general, be *any* binary relation R on $\mathbf{W}$. This elicits the question whether *any* preferential entailment relation $\mid\!\sim$ can be construed as modal, by constructing an apt accessibility relation R from the preference order $\leq$ and then formulating a modal sentence $P(X)$ which, with semantics relative to R, describes the preferred models (relative to $\leq$) of premise X:

$$\mathrm{Mod}(P(X)) = \mathrm{MaxMod}(X).$$

Unlike in the previous sections, where $P(X)$ was defined semantically as a set of interpretations, here we want a syntactic representation of X which matches a given semantic construction. The answer is yes, we can do that, even though we may not find a $P(X)$ of the simple form $\Box X$.

Given a preference order $\leq$, which we assume to be at least a preorder (reflexive transitive relation), take the accessibility relation R to be $<$, the strict partial order (irreflexive transitive relation) corresponding to $\leq$:

$$(\forall w)(\forall w')[w < w' \text{ iff } w \leq w' \text{ and } w' \not\leq w].$$

Then define

$$P(X) := X \wedge \Box\neg X,$$

which is logically stronger than X. According to the semantics induced by R, the sentence $\Box\neg X$ is true in world w if and only if X is false in all w' such that $w < w'$. Hence $P(X)$ is true in all the maximal models of X, and false in all other worlds (non-models of X as well as non-maximal models of X):

$$\mathrm{Mod}(P(X)) = \mathrm{MaxMod}(X); \text{ and } X \mid\!\sim Y \text{ iff } P(X) \models Y.$$

By the way, one tends to think intuitively that $\Box\neg X$ contradicts X, but the irreflexivity of $<$ prevents this calamity.

For the contrapositive $\mid\!\sim^*$ of $\mid\!\sim$ we see that

$$\begin{aligned} X \mid\!\sim^* Y \quad &\text{iff} \quad \neg Y \mid\!\sim \neg X \\ &\text{iff} \quad \neg Y \wedge \Box Y \models \neg X \\ &\text{iff} \quad X \models Q(Y), \end{aligned}$$

where

$$Q(Y) := \Box Y \to Y.$$

$Q(Y)$ is not in general a tautology, since $<$ is irreflexive, but $Q(Y)$ is logically weaker than Y.

We have then demonstrated that defeasible entailments $X \mid\!\sim Y$ and $X \mid\!\sim^* Y$ based on preference are respectively nonmonotonic and monotonic, and equivalent to entailments $P(X) \models Y$ and $X \models Q(Y)$, with $P(X) = X \wedge \Box\neg X$ and $Q(Y) = Y \vee \neg\Box Y$ in the modal language with semantics induced by that accessibility relation which is the strict variant of the preference order. Remember that in the non-modal language for preferential logic in Section 2 the filter $\{Y \mid X \mid\!\sim Y\}$ need not be principal, since $\mathrm{MaxMod}(X)$ need not be axiomatizable by a single sentence; similarly the ideal $\{X \mid X \mid\!\sim^* Y\}$ need not be principal. However, in the more expressive modal language of this section the corresponding filter and ideal are principal, generated respectively by the single modal sentences $P(X)$ and $Q(Y)$.

5 Proof theory

Having characterized preferential entailment modally, we now turn to the axiomatization of the accessibility relation $<$ which was used to give an appropriate semantics to the sentence $P(X) = X \wedge \Box\neg X$. In its most general form, $<$ is a Noetherian strict partial order.

Gödel-Löb logic **GL** (an important provability logic) is obtained from the minimal modal logic **K** by adding the transitivity and Löb axioms (see, for example, [6]):

$$\mathbf{GL} := \mathbf{K} \ \oplus\ \Box X \to \Box\Box X \ \oplus\ \Box(\Box X \to X) \to \Box X.$$

Segerberg [12] proved that **GL** is determined by the class of all Noetherian strict partial orders. This makes **GL** the appropriate logic to reason syntactically about modal preferential entailment.

Should the given preference order be a modular partial order or a total preorder, we need to consider the axiomatization of modularity or connectedness. In fact, we only need to consider modularity, as any total preorder can be converted to an associated modular partial order via its associated strict partial order: Let preorders S and T be *order-equivalent* iff they have the same associated strict partial orders. It is then not difficult to show that, for any modular partial order S, there is a total preorder T such that S and T are order-equivalent, and conversely, for every total preorder T

there is a modular partial order S such that S and T are order-equivalent. So one can move without loss of information from a total preorder to a modular partial order via the modal representation of information in terms of a strict accessibility relation.

Linearity, and trichotomy, cannot be axiomatized without resorting to a tense logic (i.e. a bidirectional frame with corresponding modal operators), and we conjecture that neither can modularity. But all is not lost. Just as linearity can be weakened to prohibit branching to the right (see, for example, [3, p.193]), one can weaken modularity to prohibit upward-branching over more than one level: Let an order relation be *weakly modular* if

$$(\forall w)(\forall u)(\forall v)[\text{If } w < v \text{ and } w < u \text{ then } v < u \text{ or } u < v \text{ or } \uparrow v = \uparrow u],$$

where $\uparrow v = \{v' \mid v < v'\}$ is the strict upclosure of v. The axiomatization of weak modularity is then as follows:

LEMMA 1. Let $(\mathbf{W}, <)$ be a **GL**-frame (i.e. a Noetherian strict partial order). $(\mathbf{W}, <)$ is weakly modular if and only if any sentence of the form

$$\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$$

is valid in the frame.

Proof. Left to right: Suppose $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$ is false in world w. Then

(i) $\exists u$ such that $w < u$ and $\Box X \to Y$ is false in u, i.e. $\Box X$ is true in u and Y is false in u, and

(ii) $\exists v$ such that $w < v$ and $\Box Y \to \Box X$ is false in v, i.e. $\Box Y$ is true in v and $\Box X$ is false in v.

Suppose also that $<$ is weakly modular. Then $v < u$ or $u < v$ or $\uparrow v = \uparrow u$. If $v < u$, then $\Box Y$ is true in v, so Y is true in u, contradicting (i). If $u < v$, then $\Box X$ is true in u, so $\Box X$ is true in v, contradicting (ii). If $\uparrow v = \uparrow u$, then $\Box X$ is true in u, so $\Box X$ is true in v, contradicting (ii). Therefore weak modularity of the frame implies validity of $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$.

Conversely, suppose there exist worlds u, v, w such that $w < v$ and $w < u$ and not($v < u$ or $u < v$ or $\uparrow v = \uparrow u$). Without loss of generality, we can assume that $\exists z$ such that $v < z$ and $u \not< z$.

Let Y be any sentence true in $\uparrow v$ and false in u, and let X be any sentence true in $\uparrow u$ and false in z. We see that $\Box X$ is true in u, Y is false in u, $\Box Y$ is true in v and $\Box X$ is false in v. Therefore $\Box X \to Y$ is false in u and $\Box Y \to \Box X$ is false in v. So $\Box(\Box X \to Y) \vee \Box(\Box Y \to \Box X)$ is not valid. Therefore validity of the axiom implies weak modularity. ■

Let *Modular* ***GL*** be the logic obtained from **GL** by adding the weak modularity axiom. We therefore have the following result:

THEOREM 1. *Modular* ***GL*** *is determined by the class of all Noetherian weakly modular strict partial orders.*

Nonmonotonic logics, and preferential logics in particular, do not have a well-developed theory of effective decision procedures at their disposal. Research along these lines has mostly been centered around Gentzen-style inference rules, recursively generating the set of valid entailment pairs [10, 8]. In contrast, there are many effective proof (and decision) procedures for modal logics resulting from an increased awareness of the range and significance of computational applications of modal logics, for example in applications of description logics [2]. Algorithms to determine modal satisfiability include both specialized modal satisfiability algorithms, for example tableau-based methods [7], and translation-based methods [11].

From the perspective of determining modal satisfiability, the first-order semantic characterization of modularity is of greater interest than its (weak) modal axiomatization – in modal tableaux the properties of the accessibility relation are built into the proof rules, and in first-order translations the accessibility relation is axiomatized in first-order logic.

Casting preferential reasoning as a modal satisfiability problem opens up the well-developed decision procedures for modal and description logics for use in preferential logics.

BIBLIOGRAPHY

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[3] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2002.

[4] F.M. Donini, M. Lenzerini, D. Nardi, A. Schaerf, and W. Nutt. An epistemic operator for description logics. *Artificial Intelligence*, 100(1-2):225–274, 1998.

[5] P.A. Flach and A.C. Kakas, editors. *Abduction and Induction: Essays on their Relation and Integration*. Kluwer Academic Publishers, Dordrecht, 2000.

[6] D.M. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-Dimensional Modal Logics: Theory and Applications*, volume 148 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, Amsterdam, 2003.

[7] R. Gore. Tableau methods for modal and temporal logics. In M. D'Agostino, D.M. Gabbay, R. Hähnle, and J. Posegga, editors, *Handbook of Tableau Methods*, pages 297–396. Kluwer Academic Publishers, Dordrecht, 1999.

[8] Y. Kaluzhny and D. Lehmann. Deductive nonmonotonic inference operations: antitonic representation. *Journal of Logic and Computation*, 5(1):111–122, 1995.

[9] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[10] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[11] H. De Nivelle, R. Schmidt, and U. Hustadt. Resolution-based methods for modal logics. *Logic Journal of the IGPL*, 8(3):265–292, 2000.

[12] K. Segerberg. *An essay in classical modal logic*, volume 13 of *Philosophical Studies*. 1971. PhD dissertion, Stanford University.

[13] Y. Shoham. *Reasoning about change: Time and causation from the standpoint of artificial intelligence*. The MIT Press, Cambridge, Massachusetts, 1988.

Katarina Britz

School of Computing, University of South Africa, P.O. Box 392, 0003 Pretoria, South Africa.

britzk@unisa.ac.za

Johannes Heidema

Department of Mathematical Sciences, University of South Africa, P.O. Box 392, 0003 Pretoria, South Africa.

heidej@unisa.ac.za

Willem Labuschagne

Department of Computer Science, University of Otago, P.O. Box 56, Dunedin, New Zealand.

willem@cs.otago.ac.nz