Simple Conditionals with Constrained Right Weakening

Giovanni Casini[1] , **Thomas Meyer**[2] , **Ivan Varzinczak**[3]

[1]CSC, University of Luxembourg, Luxembourg
[2]CAIR and Univ. of Cape Town, South Africa
[3]CRIL, Univ. Artois & CNRS, France
giovanni.casini@uni.lu, tmeyer@cs.uct.ac.za, varzinczak@cril.fr

Abstract

In this paper we introduce and investigate a very basic semantics for conditionals that can be used to define a broad class of conditional reasoning systems. We show that it encompasses the most popular kinds of conditional reasoning developed in logic-based KR. It turns out that the semantics we propose is appropriate for a structural analysis of those conditionals that do not satisfy the property of Right Weakening. We show that it can be used for the further development of an analysis of the notion of relevance in conditional reasoning.

1 Introduction

The structure "if C, then D" is omnipresent in communication and cognition. It is generally considered as one of the building blocks of reasoning and its formalisation in logics. An *if-then* connection can be of many different types: "For every triple of natural numbers x, y, z, if $x > y$ and $y > z$, then $x > z$", "Birds usually fly", "If you have had alcohol, you should not drive", "If you throw a stone against that window, then you will break it", "If you go out in this weather, you will probably get a cold", "If I were you, I wouldn't do that". From these examples it is clear that a conditional connection between propositions can be associated with different modalities. This includes necessity, presumption, deontic, causal, probabilistic, and counterfactual.

In this paper we investigate a semantic characterisation of a basic and general class of conditionals. For such a class to be viewed as sufficiently general, it needs to include the existing popular classes of conditionals that have been proposed. We use choice functions ranging over possible worlds, a very general formal tool that has been used for characterising conditional reasoning [Rott, 2001]. It turns out that the framework we propose is also appropriate for modelling some constraints related to the problem of *relevance*. More specifically, it provides the basis for a structural analysis of conditionals that do not satisfy the property of Right Weakening.

The basic idea of a conditional is very simple: $C \Rightarrow D$ is read as "if the *condition* C holds, then the *effect* D holds with a certain *modality*". An important question in the area of Knowledge Representation (and the related field of Philosophical Logic) is the study of appropriate formalisations of the different kinds of modality necessary for the characterisation of important reasoning behaviours.

In order to model conditional reasoning, there is usually a focus on two main aspects that need to be formalised:

1. The existence (or non-existence) of a connection between a specific type of *condition* and a specific type of *effect* (both represented by propositions).

2. The *modality* of such a connection (necessity, presumption, probability, preference, . . .).

But there is also a third aspect that is frequently neglected in this context:

3. The *relevance* of such a connection.

The importance of *relevance* in conditional reasoning has not gone unnoticed [Delgrande, 2011]. However what 'relevance' means in the context of conditional reasoning remains vague. Before proceeding with the technical aspects of our contribution, we first provide an intuitive sense of our use of the term 'relevance' in the present context.

Most of the formalisations of conditional reasoning have been built using a possible-worlds semantics, referring more or less directly to classical modal operators. This approach has made it possible to define logical systems accounting for various kinds of non-classical reasoning in surprisingly efficient ways. On the other hand, relying on possible worlds, and consequently on closed logical theories, enforces some properties (e.g. logical omniscience) that can be in conflict with some modelling goals. One limit is that the possible-worlds approach to the formalisation of conditionals ("if a *condition* C holds, then an *effect* D holds with a certain *modality*") tries to account for the degree of (epistemic or factual) necessity which is associated with the truth of D given the truth of C, but it does not account for whether the truth of D given the truth of C has any relevance for the kind of reasoning we are considering. Examples from deontic or causal reasoning can illustrate the point.

• "If you are involved in a car accident, you should remain on the spot" is an acceptable norm, but "If you are involved in a car accident, you should remain on the spot or paint yourself in blue" is not as acceptable as the former.

• "If you turn the wheel of a moving car, the car will move in a circle" is meaningful, while "If you turn the wheel of a moving car, the car will move" is not really that meaningful.

- Similarly, "If you throw a stone against the window, it will break" is meaningful, but "If you throw a stone against the window, it will break or Ann will drink tea" is not.

For all these examples, if we use a possible-worlds semantics to model the conditionals, the truth of the first conditional implies the truth of the second one (see the following section). But, even though this can be reasonable from a purely truth-theoretical point of view, if we want to determine what our obligations are in case of car accidents, or what the effects of turning the car wheel are, or those of throwing a stone, the second conditional in each example has to be considered irrelevant, if not misleading.

The semantics we propose still uses possible worlds, but allows us to define reasoning patterns that do not enforce the kind of problematic derivations indicated above.

The paper is organised as follows. In Section 2 we investigate the problem more formally and propose a first semantic construction. Section 3 shows how to take back popular properties of conditional reasoning, while Section 4 introduces an even more basic form of conditional reasoning. In Section 5 we investigate different forms of closures of conditional bases. Related Work and Conclusions follow.

2 The Formal Framework

We use a simple conditional language, containing only conditionals of the form $C \Rightarrow D$ without the possibility of nesting the conditionals or combining them via propositional operators. This is in line with the so-called KLM approach [Kraus *et al.*, 1990] or the I/O logics [Makinson and van der Torre, 2000]. This choice of language provides for a framework that combines simplicity and generality, and allows for extending the work presented here to richer conditional languages. We use a finitely generated propositional language $\mathcal{L}$, with the operators $\neg, \vee, \wedge, \rightarrow, \equiv, \bot$ having their usual meaning. Capital letters $C, D, \ldots$ will be used to refer to the propositions. Cn indicates classical propositional consequence, with $\vdash$ being the corresponding consequence relation. $\mathcal{W}$ will be used to denote the set of valuations (possible worlds) generated by $\mathcal{L}$, while $[C]_{\mathcal{W}}$ is the set of valuations in $\mathcal{W}$ satisfying a formula C (we omit the subscript whenever the context allows for it). $\hat{V}$ indicates a (canonical) formula characterising a set of valuations V. W.l.o.g., $\mathcal{W}$ could be a subset of the propositional valuations generated by $\mathcal{L}$, and Cn and $\vdash$ the supraclassical monotonic consequence operation and relation generated by $\mathcal{W}$. Our language will be $\mathcal{L}_{\Rightarrow}$, the conditional language built on top of $\mathcal{L}$: $\mathcal{L}_{\Rightarrow} \equiv_{\text{def}} \{C \Rightarrow D \mid C, D \in \mathcal{L}\}$.

As indicated in Section 1 we define a possible-worlds semantics, using choice functions as the constructions within the framework. The essential information conveyed by a conditional "If C, then D" is "if the *condition* C holds, then the *effect* D holds (with a certain modality)". Hence, the basic construction of a conditional system will simply associate to each condition the set of its effects.

Definition 1 An effect function *is a function*

$$f_+ : 2^{\mathcal{W}} \longrightarrow 2^{(2^{\mathcal{W}})}.$$

A standard conditional interpretation *is a pair*

$$\mathcal{I} = (\mathcal{W}, f_+).$$

$\mathcal{I}$ +*-satisfies a conditional* $C \Rightarrow D$ ($\mathcal{I} \Vdash_+ C \Rightarrow D$) *if there is a* $V \in f_+([C])$ *s.t.* $V \subseteq [D]$.

This function associates with each *condition* (a propositional theory, that is, a set of valuations) a set of *effects* (a set of propositional theories, that is, a set of sets of valuations). For example, a set of conditionals $\{C \Rightarrow D, C \Rightarrow E, F \Rightarrow G\}$ would be +-satisfied by an interpretation $\mathcal{I} = (\mathcal{W}, f_+)$ s.t. $f_+([C]) = \{[D], [E]\}$, $f_+([F]) = \{[G]\}$ and $f_+([H]) = \emptyset$ for every H s.t. $\nvdash (H \equiv C) \vee (H \equiv F)$.

This brings us to our first representation result. It is easy to prove that the class of the standard conditional interpretations can be used to characterise the set of conditionals closed under two basic structural properties, *Left Logical Equivalence* and *Right Weakening*:

$$\frac{\vdash C \equiv D, \quad C \Rightarrow E}{D \Rightarrow E} \quad \text{(LLE);} \qquad \frac{C \Rightarrow D, \quad \vdash D \rightarrow E}{C \Rightarrow E} \quad \text{(RW)}$$

Theorem 1 *A set of conditionals* $\mathcal{C}$ *is closed under (LLE) and (RW) iff there is a standard interpretation* $\mathcal{I} = (\mathcal{W}, f_+)$ *s.t.* $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_+ C \Rightarrow D\}$.

Sketch of proof. We omit the proof from right to left. From left to right: let $\mathcal{C}$ be a set of conditionals closed under (LLE) and (RW). We need to define an interpretation $\mathcal{I} = (\mathcal{W}, f_+)$ s.t. $\mathcal{I} \Vdash_+ C \Rightarrow D$ iff $C \Rightarrow D \in \mathcal{C}$. Let $\mathfrak{A}_{\mathcal{C}} \equiv_{\text{def}} \{C \mid$ there is a D s.t. $C \Rightarrow D \in \mathcal{C}\}$ (the set of all conditions appearing in $\mathcal{C}$) and $\mathfrak{E}_{\mathcal{C}}(C) \equiv_{\text{def}} \{D \mid C \Rightarrow D \in \mathcal{C}\}$ (the effects of a condition C according to $\mathcal{C}$). For every propositional formula C do the following:
- If $C \notin \mathfrak{A}_{\mathcal{C}}$, set $f_+(C) = \emptyset$.
- If $C \in \mathfrak{A}_{\mathcal{C}}$, then set the functions as follows:

Let $\geq_{\vdash}$ be the partial order on $\mathcal{L}$ determined by logical strength, that is, $C \geq_{\vdash} D$ iff $C \vdash D$, while $M_C \equiv_{\text{def}} \max_{>_{\vdash}}(\mathfrak{E}_{\mathcal{C}}(C))$ (the set of the logically strongest formulas in $\mathfrak{E}_{\mathcal{C}}(C)$). Set $f_+([C]) = \{[D] \mid D \in M_C\}$.

Next easy steps: prove that for every conditional $E \Rightarrow F$, $\mathcal{I} \Vdash E \Rightarrow F$ iff $E \Rightarrow F$ can be derived from (LLE) and (RW) from some $C \Rightarrow D$ s.t. $C \in \mathfrak{A}_{\mathcal{C}}$ and $D \in M_C$. Then prove that the same holds for all the $E \Rightarrow F \in \mathcal{C}$. □

This class of conditionals is quite broad, and imposes a minimal number of constraints. It represents the idea that a conditional simply associates a condition with a set of effects, and we check what holds in the valuations satisfying such effects. By using this as a starting point, and then adding constraints to the choice functions, we get semantic characterisations of many interesting classes of conditionals that have been studied in the literature (see Section 3). But (RW) holds even in this broad, basic, class of conditionals.

The counter-intuitive nature of the conditionals in the examples in Section 1 are due precisely to the validity of (RW). The statement 'remain on the spot or paint yourself in blue' is a logical consequence of 'remain on the spot'. Similarly, 'moving' is a consequence of 'moving in circle', while 'the window will break or Ann will drink a tea' is a consequence of 'the window will break'. The satisfaction of (RW) is directly connected to the use of a possible-worlds framework. Because of this, some researchers advocating for alternative kinds of semantics have suggested that the possible-worlds

approach is inherently problematic since it contains the assumption that all conditionals bear a truth value [Parent and van der Torre, 2014, p.1]. On the other hand, the core idea underlying the possible-worlds approach (modelling our reasoning referring to what holds in specific sets of conceivable situations) has some important advantages: it appears intuitive, it is general and powerful, and it is a well-developed field of investigation that offers to the user a full set of axiom schemas that can be modified and extended easily. That is, it provides simple and useful investigative tools. So, should we abandon the possible-worlds paradigm when it comes to conditional reasoning? Our response to this question is that the better course of action is to investigate ways of fitting very basic forms of conditional reasoning, possibly not satisfying (RW), into the formal possible-worlds framework.

Interestingly enough, constraining (RW) is analogous to the placing of constraints on the property of Monotonicity. In both cases we are dealing with constraints on the transitivity of the *if-then* connections:

$$\frac{\vdash C\rightarrow D,\quad D\Rightarrow E}{C\Rightarrow E}\quad \text{(Mon)}; \qquad \frac{C\Rightarrow D,\quad \vdash D\rightarrow E}{C\Rightarrow E}\quad \text{(RW)}$$

When constraining (Mon), we deal with conditional transitivity on the *condition* (on the left). In contrast, when constraining (RW) we deal with conditional transitivity on the *effect* (on the right). In non-monotonic reasoning the most interesting systems do not aim at arbitrarily invalidating Monotonicity. Instead, a rational agent should reason assuming that Monotonicity holds, and only the explicit knowledge that some situation is exceptional should prevent it from reasoning monotonically. For example, if I know that birds typically fly I conclude that robins, seagulls, and eagles fly, since I am not informed of anything to the contrary. Moreover, only the explicit knowledge that penguins do not fly prevents me from reasoning monotonically about the flying abilities of this subclass of birds [Lehmann and Magidor, 1992; Lehmann, 1995; Pearl, 1990]. The same principle can be applied to Right Weakening. (RW) is a natural reasoning pattern, and an agent should persistently apply it unless it is explicitly aware of the fact that the derived conditional would be not relevant. In order to model *irrelevance* we introduce a second function.

Definition 2 *A* relevance function *is a function*

$$f_{-} : 2^{\mathcal{W}} \longrightarrow 2^{(2^{\mathcal{W}})}.$$

A conditional interpretation *is a triple*

$$\mathcal{I} = (\mathcal{W}, f_{+}, f_{-}).$$

$\mathcal{I}$ $\pm$-satisfies *a conditional* $C \Rightarrow D$ ($\mathcal{I} \Vdash_{\pm} C \Rightarrow D$) if

- *there is a* $V \in f_{+}(C)$ *s.t.* $V \subseteq [D]_{\mathcal{W}}$;
- *for every* $V' \in f_{-}(C)$, $V' \not\subseteq [D]_{\mathcal{W}}$.

The relevance function f_{-} associates with a condition C a set of formulas that is considered to be irrelevant w.r.t. C. Informally, this new notion expresses the intuition that we apply (RW) if we are not explicitly aware that it is irrelevant. For example, if $f_{+}([C]) = \{[D]\}$ and $f_{-}([C]) = \{[(D \vee E \vee F)]\}$, then the interpretation $\pm$-satisfies $C \Rightarrow D$, $C \Rightarrow (D \vee E)$, and $C \Rightarrow (D \vee F)$, but not $C \Rightarrow (D \vee E \vee F)$ or $C \Rightarrow (D \vee E \vee F \vee G)$, since $(D \vee E \vee F)$ is an irrelevant formula and $(D \vee E \vee F \vee G)$ is the consequence of an irrelevant formula.

As shown in Figure 1, the set of $\pm$-satisfied conditionals is $\{C \Rightarrow H \mid H \in Cn(D) \setminus Cn(D \vee E \vee F)\}$.
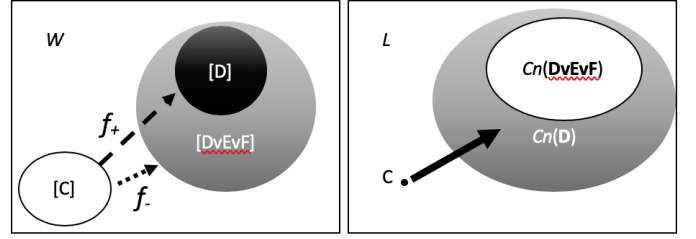

Figure 1: A conditional interpretation $\mathcal{I} = (\mathcal{W}, f_{+}, f_{-})$ s.t. $f_{+}([C]) = \{[D]\}$ and $f_{-}([C]) = \{[(D \vee E \vee F)]\}$. It $\pm$-satisfies all the conditionals $C \Rightarrow H$ s.t. H is in the grey area on the right.

Clearly the satisfaction relation $\Vdash_{\pm}$ does not satisfy (RW), but the property of *Right Logical Equivalence* continues to hold, though.

$$\frac{C\Rightarrow D,\quad \vdash D\equiv E}{C\Rightarrow E}\quad \text{(RLE)}$$

Important, (RW) is not abandoned completely: it is replaced with the weaker property of *Anti Right Weakening*:

$$\frac{C\Rightarrow D,\ \vdash D\rightarrow E,\ \vdash E\rightarrow F,\ C\not\Rightarrow E}{C\not\Rightarrow F}\quad \text{(AntiRW)}$$

or, equivalently:

$$\frac{C\Rightarrow D,\ \vdash D\rightarrow E,\ \vdash E\rightarrow F,\ C\Rightarrow F}{C\Rightarrow E}\quad \text{(AntiRW*)}$$

(AntiRW) states that, once (RW) is blocked for a formula, it is blocked for all the implied formulas. For example, suppose that "If you throw a stone at the window, it will break" is satisfied by an interpretation while "If you throw a stone at the window, it will break or Ann will drink a tea" is not. Then "If you throw a stone at the window, it will break or Ann will drink a tea or the sun will rise tomorrow" cannot be satisfied either.

(AntiRW) implies (RLE): just set F as D in the (AntiRW*) schema. Hence every set of conditionals closed under (AntiRW) is closed also under (RLE). We now obtain our second representation result: using the satisfaction relation $\Vdash_{\pm}$, the class of conditional interpretations characterises the class of sets of conditionals closed under (LLE) and (AntiRW).

Theorem 2 *A set of conditionals $\mathcal{C}$ is closed under (LLE) and (AntiRW) iff there is a conditional interpretation $\mathcal{I} = (\mathcal{W}, f_{+}, f_{-})$ s.t. $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_{\pm} C \Rightarrow D\}$.*

Sketch of proof. From right to left, we need to prove that (AntiRW) holds. Assume $\vdash D \rightarrow E, \vdash E \rightarrow F, \mathcal{I} \Vdash_{\pm} C \Rightarrow D$ and $\mathcal{I} \not\Vdash_{\pm} C \Rightarrow E$. Then there is a $V \in f_{+}([C])$ s.t. $V \subseteq [D] \subseteq [E] \subseteq [F]$. This, together with $\mathcal{I} \not\Vdash_{\pm} C \Rightarrow E$, implies that there is a $V' \in f_{-}([C])$ s.t. $V' \subseteq [E]$, that in turn implies $V' \subseteq [F]$. $V \subseteq [F]$ and $V' \subseteq [F]$ imply $\mathcal{I} \not\Vdash_{\pm} C \Rightarrow F$.

From left to right: let $\mathcal{C}$ be closed under (LLE) and (AntiRW). We define $\mathcal{I} = (\mathcal{W}, f_{+}, f_{-})$ as follows.

Let $\mathfrak{A}_{\mathcal{C}}$, $\mathfrak{E}_{\mathcal{C}}(C)$, $\geq_{\vdash}$, and M_C be defined as for Theorem 1. For every propositional formula C do the following:

- If $C \notin \mathfrak{A}_{\mathcal{C}}$, let $f_{+}([C]) = \emptyset$.
- If $C \in \mathfrak{A}_{\mathcal{C}}$, then set the functions as follows. Let $R_C = \{Cn(D) \mid D \in M_C\}$ and $f_{+}([C]) = \{[D] \mid D \in M_C\}$

(R_C is the smallest set of theories containing $\mathfrak{E}_\mathcal{C}(C)$). Let $S_C = \bigcup R_C \setminus \mathfrak{E}_\mathcal{C}(C)$ (S_C is the set of all formulas that are satisfied by the theories in $f_+(C)$ but are not relevant). Let $\max_{\geq_\vdash}(S_C)$ be the set of the strongest logical formulas in S_C, and set $f_-(C) = \{[D]_\mathcal{W} \mid D \in \max_{\geq_\vdash}(S)\}$.

Next easy step to prove: since $\mathcal{C}$ is closed under (AntiRW), $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ is a characteristic model of the set of conditionals $\mathcal{C}$. $\qquad\square$

3 Richer Conditionals

The satisfaction relation $\Vdash_\pm$ over conditional interpretations is general enough to be able to accommodate many well-known forms of conditional reasoning. In fact, many such forms of conditional reasoning can be characterised by imposing further constraints on the functions f_+ and f_-. Here we examine the introduction of a few important properties.

First, in order to characterise the closure under (LLE) and (RW), it is sufficient to set that for every formula C,

$$f_-([C]) = \emptyset \qquad (1)$$

Theorem 3 *A set of conditionals $\mathcal{C}$ is closed under (LLE) and (RW) iff there is a conditional interpretation $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ satisfying Constraint (1) s.t. $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_\pm C \Rightarrow D\}$.*

We omit the proof, since it is immediate to see that there is a full correspondence between the class of conditional interpretations satisfying Constraint (1) and the class of standard interpretations. Hence Theorem 1 applies.

Next, the property of *Conjunction on the Right*,

$$\frac{C \Rightarrow D, \quad C \Rightarrow E}{C \Rightarrow (D \wedge E)} \quad \text{(And)},$$

can be obtained by imposing that we have at most one theory describing the effects of a condition. That is, for every formula C,

$$\#(f_+([C])) \leq 1 \qquad (2)$$

Theorem 4 *A set of conditionals $\mathcal{C}$ is closed under (LLE), (AntiRW) and (And) iff there is a conditional interpretation $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ satisfying Constraint (2) s.t. $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_\pm C \Rightarrow D\}$.*

Sketch of proof. From right to left. If $\mathcal{I} \Vdash_\pm C \Rightarrow D$ and $\mathcal{I} \Vdash_\pm C \Rightarrow E$, then the only set $V \in f_+([C])$ is such that $V \subseteq [D]$, $V \subseteq [E]$ and there is no theory $V' \in f_-([C])$ s.t. $V' \subseteq [D]$ or $V' \subseteq [E]$. That implies that $V \subseteq [D] \cap [E]$ and there is no theory $V' \in f_-([C])$ s.t. $V' \subseteq [D] \cap [E]$, that is, $\mathcal{I} \Vdash_\pm C \Rightarrow (D \wedge E)$.

From left to right. Let $\mathcal{C}$ be closed under (LLE), (AntiRW) and (And), and let $\mathfrak{A}_\mathcal{C}$, $\mathfrak{E}_\mathcal{C}(C)$, and M_C be defined as for Theorem 1. Due to (And), for every $C \in \mathfrak{A}_\mathcal{C}$ there is only one formula D in M_C (modulo logical equivalence.). Using the same construction used for Theorem 2 we can build the characteristic model. $\qquad\square$

Next, we observe that for *consistency preservation*,

$$\frac{\nvdash \neg C}{C \not\Rightarrow \bot} \quad \text{(Cons)}$$

it is sufficient to impose that

$$\text{If } [C] \neq \emptyset, \text{ then } \emptyset \notin f_+([C]). \qquad (3)$$

Continuing, we observe that for Reflexivity ($C \Rightarrow C$) to be satisfied, we simply need to impose that for every C,

$$\begin{aligned} &\text{there is a } V \in f_+([C]) \text{ s.t. } V \subseteq [C]; \\ &\text{there is no } V \in f_-([C]) \text{ s.t. } V \subseteq [C]. \end{aligned} \qquad (4)$$

And for Monotonicity (Mon) to hold, it is sufficient to impose that for every C, D s.t. $\vdash C \rightarrow D$,

$$\begin{aligned} &\text{for every } V \in f_+([D]) \text{ there is a } V' \in f_+([C]) \text{ s.t. } V' \subseteq V; \\ &\text{for every } V \in f_-([C]) \text{ there is a } V' \in f_-([D]) \text{ s.t. } V' \subseteq V. \end{aligned} \qquad (5)$$

Theorem 5 *A set of conditionals $\mathcal{C}$ is closed under (LLE), (AntiRW) and (Mon) iff there is a conditional interpretation $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ satisfying Constraint (5) s.t. $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_\pm C \Rightarrow D\}$.*

Sketch of proof. From right to left the proof is straightforward. From left to right, let $\mathcal{C}$ be closed under (LLE), (AntiRW) and (Mon), and let $\mathfrak{A}_\mathcal{C}$, $\mathfrak{E}_\mathcal{C}(C)$, and M_C and S_C be defined as for Theorems 1 and 2. Let $\mathcal{I}$ be defined as in the proof of Theorem 2. $\mathcal{I}$ is a characteristic model of $\mathcal{C}$. We need to prove that $\mathcal{I}$ satisfies Constraint 5. Let $\vdash C \rightarrow D$. To have (Mon) it is necessary that for every formula E in M_D there is a formula F in M_C s.t. $\vdash F \rightarrow E$, and this enforces the first part of Constraint (5). For the second part of the constraint, assume it is not satisfied, that is, assume that for some $V \in f_-(C)$ there is no $V' \in f_-(D)$ s.t. $V' \subseteq V$. According to the construction of $\mathcal{I}$, this implies that there is a formula H in $\max_{\geq_\vdash}(S_C)$ that is not implied by any formula I in $\max_{\geq_\vdash}(S_D)$ which, in turn, implies that $\mathcal{I} \Vdash_\pm D \Rightarrow H$ while $\mathcal{I} \nVdash_\pm C \Rightarrow H$, contradicting (Mon). $\qquad\square$

If the properties (And) and (RW) are satisfied, then to add (Mon) we just require that if $\vdash C \rightarrow D$, then $V \subseteq V'$, with $f_+(C) = \{V\}$ and $f_+(D) = \{V'\}$. Finally, assuming (And) and (RW), the question of how to impose the satisfaction of relevant properties has already been thoroughly investigated by Rott [2001, Chapter 7], and we will investigate in future work how to enforce such properties in our framework.

4 Dropping Anti Right Weakening

Is (AntiRW) a reasonable property? It certainly makes sense for some versions of conditional reasoning. A good example is a basic commonsense notion of causality. Given some context, C causes D ($C \Rightarrow D$) iff D would be true in case C were true, but it would not necessarily hold in the absence of C. This could be modelled in the present framework as follows: $f_+([C])$ corresponds to the set of situations associated with the occurrence of C, while $f_-([C])$ corresponds to the set of situations associated with the non-occurrence of C. For example, assume that our background theory contains some basic astronomical notions, such as that we consider it impossible that the sun will not rise tomorrow ($\vdash sun\_tomorrow$). Also we know that the passage of the moon in front of the sun will cause an eclipse phenomenon. That is, $passage\_moon \Rightarrow eclipse$. We can define, for example, a conditional interpretation $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ s.t. $\mathcal{W} \subseteq [sun\_tomorrow]$ (since $\vdash sun\_tomorrow$), $f_+([passage\_moon]) = \{\mathcal{V}\}$ s.t. $\mathcal{V} \subseteq [eclipse]$ and $f_-([passage\_moon]) = \{\mathcal{V}'\}$ s.t.

$\mathcal{V}' \subseteq [\neg eclipse]$, with both $\mathcal{V}$ and $\mathcal{V}'$ nonempty. By Definition 2, we have $\mathcal{I} \Vdash_{\pm} passage\_moon \Rightarrow eclipse$. However, since in every valuation the sun will rise tomorrow ($\mathcal{W} \subseteq [sun\_tomorrow]$), we have $\mathcal{V}, \mathcal{V}' \subseteq [sun\_tomorrow]$, and consequently $\mathcal{V}, \mathcal{V}' \subseteq [eclipse \vee sun\_tomorrow]$. Given $\mathcal{I}$, despite $\mathcal{V} \subseteq [eclipse \vee sun\_tomorrow]$, by Definition 2 we have $\mathcal{I} \not\Vdash_{\pm} passage\_moon \Rightarrow eclipse \vee sun\_tomorrow$, because $\mathcal{V}' \subseteq [eclipse \vee sun\_tomorrow]$ too, that is, despite we believe that the passage of the moon in front of the sun causes an eclipse ($passage\_moon \Rightarrow eclipse$), we do not believe that the passage of the moon in front of the sun causes an eclipse or that the sun will rise tomorrow ($passage\_moon \Rightarrow eclipse \vee sun\_tomorrow$). Furthermore, the same will hold for all the formulas implied by ($eclipse \vee sun\_tomorrow$). If ($eclipse \vee sun\_tomorrow$) holds independently from $passage\_moon$, then so does ($eclipse \vee sun\_tomorrow \vee Ann\_has\_tea$), and $passage\_moon \Rightarrow (eclipse \vee sun\_tomorrow \vee Ann\_has\_tea)$ should not hold. This is exactly what (AntiRW) imposes.

On the other hand, (AntiRW) seems to be too strong a constraint in other contexts. To see, consider conditional desires as an example. After her high-school diploma (hsd), Ann would like to get a college degree (cd), that is, $hsd \Rightarrow cd$. To get a college degree it is necessary to go to a college (gc), have enough money (m) to pay the tuition fees, and study hard (sh), that is, $\vdash cd \rightarrow gc \wedge m \wedge sh$. Ann would like to have money and to go to college ($hsd \Rightarrow m \wedge gc$), but she does not like the idea of studying hard, even if associated to having money ($hsd \not\Rightarrow m \wedge sh$). But for sure Ann would like to have money ($hsd \Rightarrow m$). This set of conditionals is not compatible with (AntiRW): if $hsd \Rightarrow gc$ is satisfied, but $hsd \Rightarrow m \wedge sh$ is not (where $\vdash cd \rightarrow gc \wedge m \wedge sh$), then (AntiRW) imposes that also $hsd \Rightarrow m$ cannot be satisfied.

(AntiRW) holds because the relation $\Vdash_{\pm}$ allows for the blocking of the weakening of the formulas associated with the sets of valuations in $f_+([C])$ using the formulas associated with the sets of valuations in $f_-([C])$. However, the same does not hold in the other direction. Once a theory is indicated as irrelevant by $f_-([C])$, the definition of $\Vdash_{\pm}$ does not allow us to conclude that information implied by some theory in $f_-([C])$ is relevant. One way of resolving this issue is to refine the satisfaction mechanism in $\Vdash_{\pm}$, allowing for a richer interaction between the theories in $f_+(\cdot)$ and the ones in $f_-(\cdot)$. In order to do that, we introduce a notion of *inferential closeness* between formulas based on their logical strength. Given C, D, E, we say that D is inferentially at least as close as E to C ($D \geq_C E$) if $E \vdash D$ and $D \vdash C$. For the strict relation, D is said to be *inferentially closer* than E to C ($D >_C E$) if $E \vdash D$, $D \not\vdash E$, and $D \vdash C$, that is, to infer C from E it is necessary to "pass by" the information represented by D. Semantically, $D >_C E$ corresponds to saying that $[E] \subset [D] \subseteq [C]$, as shown in Figure 2.

Using this notion of inferential closeness, we refine the interaction between $f_+(\cdot)$ and $f_-(\cdot)$ by requiring that (RW) on a formula D is blocked only in case for every formula represented in $f_+([C])$ and implying D there is a formula represented in $f_-([C])$ that is inferentially closer to D. That is, if we can derive $C \Rightarrow D$ from $C \Rightarrow E$ by (RW), such a derivation is blocked only in case we know that $C \Rightarrow F$ is
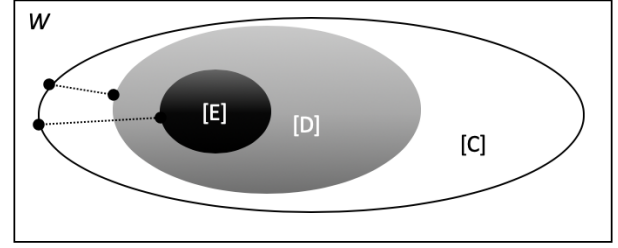

Figure 2: Both D and E imply C, but D is *inferentially closer* to C.

irrelevant, and F is inferentially closer to D than E.

Definition 3 *Let* $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ *be a* conditional interpretation. $\mathcal{I}$ ⊚-satisfies a conditional $C \Rightarrow D$ ($\mathcal{I} \Vdash_{\circledcirc} C \Rightarrow D$) if there is a $V \in f_+([C])$ s.t.

• $V \subseteq [D]_W$; and

• for every $V' \in f_-([C])$ s.t. $V' \subseteq [D]_W$, $\hat{V}' \not\geq_C \hat{V}$ (that is, $V' \not\supseteq V$).

Observe that $\Vdash_{\circledcirc}$ allows to use conditional interpretations to model (and characterise) an elementary form of conditional reasoning. This brings us to our next representation result.

Theorem 6 *A set of conditionals* $\mathcal{C}$ *is closed under (LLE) and (RLE) iff there is a conditional interpretation* $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ *s.t.* $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_{\circledcirc} C \Rightarrow D\}$.

Sketch of proof. We only consider from left to right, since the other direction is immediate. Let $\mathcal{C}$ be a set of conditionals closed under (LLE) and (RLE). Let $\mathfrak{A}_{\mathcal{C}}$ and $\mathfrak{E}_{\mathcal{C}}(C)$ be defined as for Theorem 1.

For every propositional formula C do the following:

• If $C \notin \mathfrak{A}_{\mathcal{C}}$, set $f_+([C]) = \emptyset$.

• If $C \in \mathfrak{A}_{\mathcal{C}}$, then set the functions as follows:
$f_+(C) = \{V \mid$ there is a $D \in \mathfrak{E}_{\mathcal{C}}(C)$ s.t. $\vdash D \equiv \hat{V}\}$, and
$f_-(C) = \{V \mid$ there is no $D \in \mathfrak{E}_{\mathcal{C}}(C)$ s.t. $\vdash D \equiv \hat{V}\}$.
It is easy to prove that $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ is a characteristic model of $\mathcal{C}$. $\qquad\square$

Observe further that, in order to recover (AntiRW), it is sufficient to impose the condition that for every formula C,

for every $V \in f_-([C])$ there is no $V' \in f_+([C])$ s.t. $V' \supseteq V$ (6)

Constraint (6) unlocks the next representation result.

Theorem 7 *A set of conditionals* $\mathcal{C}$ *is closed under (LLE) and (AntiRW) iff there is a conditional interpretation* $\mathcal{I} = (\mathcal{W}, f_+, f_-)$ *satisfying Constraint (6) s.t.* $\mathcal{C} = \{C \Rightarrow D \mid \mathcal{I} \Vdash_{\circledcirc} C \Rightarrow D\}$.

Sketch of proof. From right to left, we need to prove that (AntiRW) holds. Let $[D] \in f_+([C])$, and $[D \vee E] \in f_-([C])$; by Constraint (6) for every F, $\mathcal{I} \not\Vdash_{\circledcirc} C \Rightarrow (D \vee E \vee F)$. From left to right, the interpretation defined in the proof of Theorem 2 does the job. $\qquad\square$

5 Reasoning with Conditional Bases

The next important issue is an investigation of *entailment* for *conditional bases* (any finite set $\mathcal{B} = \{C_1 \Rightarrow D_1, \ldots, C_n \Rightarrow D_n\}$). Given a conditional base and the representation results we obtained, what forms of entailment are appropriate?

The obvious starting point for this investigation, and the only one we consider in this paper, is a classical Tarskian approach to entailment: $\mathcal{B} \approx C \Rightarrow D$ if $C \Rightarrow D$ is satisfied by every model of $\mathcal{B}$. Let us first consider the case of $\Vdash_\circledcirc$.

Definition 4 *Let $\mathcal{B}$ be a conditional base. $\mathcal{B} \approx_\circledcirc C \Rightarrow D$ if for every conditional interpretation $\mathcal{I}$,*

If $\mathcal{I} \Vdash_\circledcirc E \Rightarrow F$ for every $E \Rightarrow F \in \mathcal{B}$, then $\mathcal{I} \Vdash_\circledcirc C \Rightarrow D$.
$Cn_\circledcirc$ is the corresponding entailment operation:

$$Cn_\circledcirc(\mathcal{B}) = \{C \Rightarrow D \mid \mathcal{B} \approx_\circledcirc C \Rightarrow D\}.$$

What kind of reasoning is sanctioned by this form of entailment? Note first that Lemma 1 shows that the properties we are considering are closed under intersection (proof omitted).

Lemma 1 *Let $\mathcal{C}$ and $\mathcal{C}'$ be sets of conditionals closed under $(*)$, with $(*) \in \{(LLE), (RLE), (RW), (AntiRW)\}$. Also $\mathcal{C} \cap \mathcal{C}'$ is closed under $(*)$.*

$Cn_\circledcirc(\cdot)$ does a very simple thing: $Cn_\circledcirc(\mathcal{B})$ is the closure of $\mathcal{B}$ under (LLE) and (RLE).

Theorem 8 *Let $\mathcal{B}$ be a conditional base and $C \Rightarrow D$ be any conditional. $C \Rightarrow D \in Cn_\circledcirc(\mathcal{B})$ iff $C \Rightarrow D$ can be derived from some $E \Rightarrow F \in \mathcal{B}$ using (RLE) and (LLE).*

Sketch of proof. By Lemma 1, $Cn_\circledcirc(\mathcal{B})$ must be closed under (LLE) and (RLE). We use a procedure analogous to the one used for Theorem 6 to build a model of $\mathcal{B}$: it is sufficient to consider the set $\mathfrak{A}_\mathcal{B}$ and the set $\mathfrak{E}_\mathcal{B}(C)$ for every $C \in \mathfrak{A}_\mathcal{B}$, and treat the formulas modulo $\vdash$-equivalence. Such a model is the closure under (LLE) and (RLE) of $\mathcal{B}$, hence it is the smallest model of $\mathcal{B}$ (in terms of the set of satisfied conditionals), and characterises $Cn_\circledcirc(\mathcal{B})$. □

Next we consider what happens if we add (AntiRW). That is, if we refer to the satisfaction relation $\Vdash_\pm$.

Definition 5 *Let $\mathcal{B}$ be a conditional base. $\mathcal{B} \approx_\pm C \Rightarrow D$ if for every conditional interpretation $\mathcal{I}$,*

If $\mathcal{I} \Vdash_\pm E \Rightarrow F$ for every $E \Rightarrow F \in \mathcal{B}$, then $\mathcal{I} \Vdash_\pm C \Rightarrow D$.
$Cn_\pm$ is the corresponding entailment operation:

$$Cn_\pm(\mathcal{B}) = \{C \Rightarrow D \mid \mathcal{B} \approx_\pm C \Rightarrow D\}.$$

Given Lemma 1, $Cn_\pm(\mathcal{B})$ must be closed under (LLE) and (AntiRW). (AntiRW*) can be used as a derivation rule.

Theorem 9 *Let $\mathcal{B}$ be a conditional base and $C \Rightarrow D$ be any conditional. $C \Rightarrow D \in Cn_\pm(\mathcal{B})$ iff $C \Rightarrow D$ can be derived from $\mathcal{B}$ using (AntiRW*) and (LLE).*

Sketch of proof. By Lemma 1, $Cn_\circledcirc(\mathcal{B})$ must be closed under (LLE) and (AntiRW). We build a model $\mathcal{I}$ of $\mathcal{B}$ using the following procedure: Let $\mathfrak{A}_\mathcal{B} = \{C \mid \vdash C \equiv C'$ for some C' s.t. $C' \Rightarrow D \in \mathcal{B}$ for some $D\}$ and let $\mathfrak{E}_\mathcal{B}(C) = \{D \mid C \Rightarrow D \in \mathcal{B}\}$.

For every propositional formula C do the following:
- If $C \notin \mathfrak{A}_\mathcal{C}$, set $f_+(C) = \emptyset$.
- If $C \in \mathfrak{A}_\mathcal{C}$, then set the functions as follows:

Let $\geq_\vdash$ be defined as in Theorem 1, and let $m_C = \min_{>_\vdash}(\mathfrak{E}_\mathcal{C}(C))$. Set $f_+([C]) = \{[D] \mid \mathfrak{E}_\mathcal{C}(C)\}$ and $f_-([C]) = \{[E] \mid D >_\vdash E$ for some $D \in m_C\}$.

We need to prove three more easy things:

- that $\mathcal{I}$ is a model of $\mathcal{B}$ closed under (LLE) and (AntiRW);
- that $\mathcal{I}$ is the smallest (w.r.t. the set of satisfied conditionals) such model of $\mathcal{B}$;
- that $\mathcal{I} \Vdash_\pm C \Rightarrow D$ iff $C \Rightarrow D$ can be derived from $\mathcal{B}$ using (LLE) and (AntiRW*). □

We shall leave the investigation of more refined, non-Tarskian forms of entailment as future work.

6 Related Work and Conclusions

There have been many attempts to formalise the notion of relevance, but the version of relevance we consider here is quite different from that usually found in the literature. For example, there is a whole family of *Relevant Logics* [Restall, 2006], but their investigation of the notion of relevance departs from a more proof-theoretical perspective, aimed at modelling forms of entailment in which the antecedents play an active role in the derivation of the conclusion. In the area of conditional reasoning, Delgrande [2011] has stressed the importance of analysing *relevance*, and with Pelletier [1998] they investigated the notion of relevance. However, their focus is on the constraints on (Mon), rather than on (RW).

Regarding logics that do not satify (RW), an interesting example by Parent and van der Torre [2014] proposes a deontic system satisfying only (RLE), based on the semantics of I/O logics. More in line with our semantic approach is the proposal by Britz et al. [2011], but they weaken (RW) with an approach that is different from the present one, and their system satisfies different properties.

In this paper we have presented a possible-worlds semantics that allows us to avoid properties like (RW) or (And), classically associated with the possible-worlds approach. We have introduced (AntiRW), a constrained form of (RW), and we have initiated an investigation into basic forms of entailment in this framework. These are only first steps into this area, though, and much more work still needs to be done. To start off with, it is necessary to consider constrained versions of (RW) other than (AntiRW). It is also necessary to develop the investigation in Section 3 further. That is, to consider the interaction between $\Vdash_\pm$ and $\Vdash_\circledcirc$ with the other structural properties. Similarly, we have only touched the surface when it comes to the investigation into entailment in Section 5. Finally, an interesting option would be the use of negated conditionals into the language, in which we explicitly represent the case where an agent considers some conditionals as not holding. This additional expressivity would allow us to define more interesting constrained versions of (RW). It could also be combined with forms of entailment that use negated conditionals to constrain monotonicity [Booth and Paris, 1998].

References

[Booth and Paris, 1998] R. Booth and J.B. Paris. A note on the rational closure of knowledge bases with both positive and negative knowledge. *Journal of Logic, Language and Information*, 7(2):165–190, 1998.

[Britz *et al.*, 2011] K. Britz, J. Heidema, and I. Varzinczak. Constrained consequence. *Logica Universalis*, 5(2):327–350, 2011.

[Delgrande and Pelletier, 1998] J. Delgrande and F. Pelletier. A formal analysis of relevance. *Erkenntnis*, 49(2):137–173, Sep 1998.

[Delgrande, 2011] J. Delgrande. What's in a default? thoughts on the nature and role of defaults in non-monotonic reasoning. In G. Brewka, V.W. Marek, and M. Truszczynski, editors, *Nonmonotonic Reasoning. Essays Celebrating Its 30th Anniversary*. College Publications, 2011.

[Kraus *et al.*, 1990] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.*, 44:167–207, 1990.

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

[Lehmann, 1995] D. Lehmann. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence*, 15(1):61–82, 1995.

[Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29(4):383–408, 2000.

[Parent and van der Torre, 2014] X. Parent and L. van der Torre. "Sing and Dance!". In F. Cariani, D. Grossi, J. Meheus, and X. Parent, editors, *Deontic Logic and Normative Systems*, pages 149–165, Cham, 2014. Springer International Publishing.

[Pearl, 1990] J. Pearl. System Z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 1990.

[Restall, 2006] G. Restall. Relevant and substructural logics. In D.M. Gabbay and J. Woods, editors, *Handbook of the History of Logic*, volume 7: Logic and the Modalities in the Twentieth Century, pages 289–398. Elsevier North-Holland, 2006.

[Rott, 2001] H. Rott. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford University Press, 2001.