Revision by Translation

Dov Gabbay
Department of Computer Science,
King's College, London
dg@dcs.kcl.ac.uk

Odinaldo Rodrigues
Department of Computer Science,
King's College, London
rodrigu@dcs.kcl.ac.uk

Alessandra Russo
Department of Computing,
Imperial College, London
ar3@doc.ic.ac.uk

Abstract

In this paper, we show that it is possible to accomplish belief revision in any logic which is translatable to classical logic. We start with the example of the propositional modal logic K and show that a belief operation in K defined in terms of $K's$ translation to classical logic verifies the AGM postulates. We also consider the case of non-classical logics by taking Belnap's four-valued logic [5] as an example. A sound and complete axiomatization of that logic in classical logic is given and then used in the definition of a belief revision operator for the logic in terms of a classical AGM operator. We show that the operator also verifies the AGM postulates whenever they make sense in the context of non-classical logics.

1 Introduction

This paper will present a method for revision of theories in logics other than classical logic. The idea is to translate the other logic into first-order classical logic, perform the revision there and then translate back. The general schema looks as follows.

Let $*_a$ be a revision process in classical logic. Typically, given a classical logic theory Δ an input formula ψ[1], the operation $*_a$ gives us a new theory $\Gamma = \Delta *_a \psi$, corresponding to the result of the revision of Δ by ψ. Ideally, $*_a$ has some desirable properties, for instance, the well known AGM postulates for belief revision (see Section 2).

We would like to *export* this machinery to other logics. For example, given a theory Δ of some logic L and an input L-sentence ψ, can we define a revision operation $*_L$ such that $\Delta *_L \psi$ is a revised L theory and $*_L$ satisfies the AGM postulates? Can we make use of the revision operator of classical logic?

This paper presents such a method. The idea is to translate the object logic L into classical logic, perform the AGM revision there and translate the results back. Suppose that τ denotes a translation function from L into classical logic and T^τ is a classical logic theory encoding the basic properties of the logic L.

[1] There is no special need that the input is a single formula ψ. It can be a theory Ψ. The AGM postulates work for input theories as well.

If the axiomatization given by T^τ is sound and complete, we have that for all Δ and α of the logic L

$$\Delta \vdash_L \alpha \text{ iff } T^\tau \cup \Delta^\tau \vdash \alpha^\tau \tag{1}$$

Therefore, we can define a revision operator $*_L$ in the logic L as follows:

Definition 1 [Belief revision in L] Let $*_a$ be a revision operator for classical logic, and let τ, T^τ be as above. We define

$$\Delta *_L \psi = \{\beta \mid \Delta^\tau *_a (\psi^\tau \wedge T^\tau) \vdash \beta^\tau\}$$

The motivation for this definition is as follows. Δ^τ is the translation of the original L-logic theory Δ. Δ is to be revised by ψ, which in classical logic is translated as ψ^τ. We revise instead Δ^τ by ψ^τ. However, in classical logic the properties of the object logic (T^τ) have to be added as well, since it describes how the object logic works, and we want it to be preserved in the revision process (i.e. we want the resulting revised theory to satisfy T^τ), so we revise by $\psi^\tau \wedge T^\tau$[2].

Of course, the details have to be worked out. The difficulties mainly have to do with the notion of inconsistency in L. L may have theories Δ which are considered L-inconsistent while their translation Δ^τ is classically consistent[3]. Thus, we may have a situtation in L where Δ is L-consistent, the input ψ is L-consistent, but $\Delta \cup \{\psi\}$ is L-inconsistent and requires L-revision. However, when we translate into classical logic we set $\Delta^\tau \cup \{\psi^\tau\} \cup T^\tau$ and this theory is classically consistent and so classical revision will be an expansion. We therefore need to write some additional axioms say Acc(for acceptability) in classical logic that will make $\Delta^\tau \cup \{\psi^\tau\} \cup T^\tau \cup Acc$ classically inconsistent, whenever $\Delta \cup \{\psi\}$ is L-inconsistent, and thus trigger a real revision process in classical logic.

There is a problem, however, with this approach. Classical revision of $\Delta^\tau \cup \{\psi^\tau\} \cup T^\tau \cup Acc$ may give us a theory Δ_c of classical logic such that the reverse translation $\Delta_L = \{\alpha \mid \Delta_c \vdash \alpha^\tau\}$ is not a theory we are happy with in L. Put in other words, when we look at the relation between $\Delta \cup \{\psi\}$ and Δ_L we are not happy in L to consider Δ_L as the L-revision of $\Delta \cup \{\psi\}$. The reason that such a situation may arise has to do with the fact that the notion of inconsistency in classical logic is too strong. We now explain why: if K is a consistent theory in classical logic and $K \cup \{\psi\}$ is inconsistent in classical logic, then $\mathrm{Cn}(K \cup \{\psi\})$ is the set of all wffs. Our revision intuition wants to take a consistent subset K' of $\mathrm{Cn}(K \cup \{\psi\})$.

In the logic L with K_L and ψ_L and with a different notion of inconsistency, the theory $\mathrm{Cn}_L(K_L \cup \{\psi_L\})$ may not be the set of all wffs of L. We still want to

[2]Our method is restricted to logics L which have a translation τ which can be characterized by a classical theory T^τ. If the translation is, for example, semantically based, this means that the semantics of L can be expressed by a first-order theory T^τ, as is the case in many modal logics.

[3]In paraconsistent logics, for example, $p \wedge \neg p$ is considered inconsistent but we do not have $p \wedge \neg p \vdash q$. Equation (1) of the translation still holds, i.e., for all α, $p \wedge \neg p \vdash \alpha$ iff $(p \wedge \neg p)^\tau \cup T^\tau \vdash \alpha^\tau$, but in classical logic $(p \wedge \neg p)^\tau \cup T^\tau$ is consistent.

get a consistent subset K'_L of $\mathrm{Cn}_L(K_L \cup \{\psi_L\})$ as the revision. Our strategy of revision by translation may give us a revised theory via translation which is not a subset of $\mathrm{Cn}_L(K_L \cup \{\psi_L\})$ because in classical logic $\mathrm{Cn}(K^\tau_L \cup \{\psi^\tau_L\} \cup T^\tau \cup Acc)$ is too large (i.e., all wffs) and gives the revision process too much freedom. One way to solve this difficulty is to tighten up the revision process in classical logic.

The structure of the paper is as follows: in Section 2, we provide a quick introduction to the theory of Belief Revision. We analyse the meaning of these postulates for both classical and non-classical logics. This is followed in Section 3, by the application of the idea of revision by translation to the modal logic K. In Section 4 we present Belnap's four-valued logic along with a sound and complete axiomatization in classical logic. In Section 5, we analyse in more detail the application of the revision method proposed previously for non-classical logics, by taking Belnap's logic as an example. We finish the paper with some conclusions and comments in Section 6.

2 Belief Revision

The term Belief Revision is used to describe the kind of information change in which an agent reasoning about his beliefs about the world is forced to adjust them in face of new (possibly contradictory) information. One important assumption in the process is that the world is taken as a static entity. Even though changes in the world itself are not considered, the agent reasons about his knowledge about the world, which may be incorrect or incomplete. Therefore, Belief Revision is an intrinsically non-monotonic form of reasoning.

When the set of beliefs held by an agent is closed under the consequence relation of some formal language, it is usually called a *belief set*. Some variants of the standard belief revision approach also consider the case when the focus is done on a finite set of beliefs, called the *belief base*. These variants are usually called *base revision*. If Cn is the consequence relation of a given logic and K is a belief set, then it is assumed that $K = \mathrm{Cn}(K)$. Similarly, if for a belief φ and a belief set K, $\varphi \in K$, we say that φ is *accepted* in K.

The whole framework of Belief Revision is governed by some desiderata of the operations on belief sets, called *the AGM postulates for belief revision*. The term "AGM" stands for the initials of the main proposers of the theory, namely, Alchourrón, Gärdenfors and Makinson. According to the AGM theory [9], there are three main types of belief change:

- Expansion, when new information is consistent with the current belief set. All is necessary to do is to close the union of the previous belief set together with the new sentence under the consequence relation.

- Contraction, when the agent is forced to retract some beliefs. Notice that, since the belief set is closed under the consequence relation, in order to retract a belief φ, it is also necessary to remove other beliefs that imply φ.

- Revision, which is the acceptance of new information contradicting the current belief set and the subsequent process of restoring the consistency of that belief set whenever the new information is not itself contradictory.

Thus, the interesting cases are contractions and revisions. In fact, there are corresponding identities to translate between the two processes: the *Levi Identity* defines revisions in terms of contractions and the *Harper Identity* defines contractions in terms of revisions. We will concentrate on the revision part here.

The general task of the revision process is to determine what is *rational* to support after a new contradictory belief is accepted. As we mentioned before, some general postulates describe ideal properties of the operation. One of these properties is sometimes referred to as the principle of *informational economy* [9, page 49]:

> "... when we change our beliefs, we want to retain as much as possible of our old beliefs – information is not in general gratuitous, and unnecessary losses of information are therefore to be avoided."

One of the main references to the general theory of belief revision is the book "Knowledge in Flux", by Peter Gärdenfors [9]. Other references include, for instance, [2, 3, 1, 10].

We now present the postulates for the revision operation as given in [9], pages 54–56. The following conventions are assumed: K is a set of formulae of the language representing the current belief set and A (B) is a formula representing the new piece of information. We use the symbol $*_a$ to denote an AGM belief revision operator. Thus, $K *_a A$ stands for the revision of K by A. The symbol $K_\perp$ denotes the inconsistent belief set, and is equivalent to the consequences of all formulae in the language.

AGM postulates for Belief Revision (in classical logic)

(K*1) $K *_a A$ is a belief set

This postulate requires that the result of the revision operation is also a belief set. One can perceive this as the requirement that the revised set be also closed under the consequence relation.

(K*2) $A \in K *_a A$

(K*2) is known as the *success postulate* and corresponds to Dalal's *principle of primacy of the update* [6]. It basically says that the revision process should be successful in the sense that the new belief is effectively accepted in the revised belief state.

(K*3) $K *_a A \subseteq \mathrm{Cn}(K \cup \{A\})$

(K*4) If $\neg A \notin K$, then $\mathrm{Cn}(K \cup \{A\}) \subseteq K *_a A$

(K*5) $K *_a A = K_\perp$ only if A is unsatisfiable

To understand what the above three postulates (K*3)–(K*5) say, we need to consider two cases. Let $K_1 = K *_a A$.

Case 1: $K \cup \{A\}$ is consistent in classical logic.

In this case, AGM says that we want $K_1 = K*_a A$ to be equal to the closure of $K \cup \{A\}$:

- postulate (K*3) says that $K*_a A \subseteq \mathrm{Cn}(K \cup \{A\})$.
- postulate (K*4) says that $\mathrm{Cn}(K \cup \{A\}) \subseteq K*_a A$.
- postulate (K*5) is not applicable, since $K*_a A$ is consistent.

Case 2: $K \cup \{A\}$ is inconsistent.

In this case, let us see what the postulates (K*3)–(K*5) say about K_1:

- postulate (K*3) says nothing about K_1. If $K \cup \{A\}$ is classically inconsistent, then any theory whatsoever is a subset of $\mathrm{Cn}(K \cup \{A\})$, because this theory is the set of all formulae.
- postulate (K*4) says nothing. Since $K \cup \{A\}$ is inconsistent in classical logic, we have $\neg A \in K$ (since K is a closed theory), so (K*4) is satisfied, because it is an implication whose antecedent is false.
- To understand what postulate (K*5) says in our case, we distinguish two subcases:

  (2.1) A is consistent.

  (2.2) A is inconsistent.

  Postulate (K*5) says nothing about $K_1 = K*_a A$ in case (2.2) above, it however requires K_1 to be consistent, whenever A is a consistent – case (2.2).

The above case analysis shows that the AGM postulates (K*3)–(K*5) have something to say only when $K \cup \{A\}$ is consistent, or if not when A is consistent. The particular way of writing these postulates as above makes use of technical properties of classical logic (the way inconsistent theories prove everything).

When we check the AGM postulates for logics other than classical, we may have a different notion of consistency and so we are free to interpret what we want the revision to do in the case of inconsistency according to what is reasonable in the object (non-classical) logic). AGM for classical logic gives us no clue beyond (K*5) as to what to require when $(K*_a A) \cup \{B\}$ is inconsistent.

Summary of (K*3)–(K*4)

Postulates (K*3)–(K*4) effectively mean the following:

($\mathrm{K}^*_{3,4}$) If A is consistent with K, then $K*_a A = \mathrm{Cn}(K \cup \{A\})$.

If K is finite, we can take it as a formula and the postulate above corresponds to one of the rules in Katsuno and Mendelzon's rephrasing of the AGM postulates for finite knowledge bases ([11], page 187):

(R2) If $K \wedge A$ is satisfiable, then $K *_a A \leftrightarrow K \wedge A$.

For non-classical logics, where the notion of consistency is different, we need check only ($\mathrm{K}^*_{3,4}$) and (K^*5).

(K^*6) If $A \equiv B$, then $K *_a A \equiv K *_a B$

(K^*6) specifies that the revision process should be independent of the syntactic form of the sentences involved. It is called the *principle of irrelevance of syntax* by many authors, see for instance, [6].

(K^*7) $K *_a (A \wedge B) \subseteq \mathrm{Cn}((K *_a A) \cup \{B\})$

(K^*8) If $\neg B \notin K *_a A$, then $\mathrm{Cn}(K *_a A \cup \{B\}) \subseteq K *_a (A \wedge B)$

To understand what postulates (K^*7)–(K^*8) are saying, we again have to make a case analysis. The postulates have to do with the relationship of inputing (A, B) as a sequence (first revising by A, then expanding by B), as compared with revising by $\{A, B\}$ at the same time (i.e, revising by $A \wedge B$). It is well known that AGM does not say enough about sequences of revisions and their properties. These postulates are the bare minimum (see, for instance, [7, 8]).

We distinguish the following cases:

Case 1: A is consistent with K.

In this case, $K_1 = K *_a A$ is equal (by previous postulates) to $\mathrm{Cn}(K \cup \{A\})$.

(1.1) B is consistent with K_1. In this case, the antecendet of (K^*8) holds and (K^*7) and (K^*8) together effectively say that $\mathrm{Cn}((K *_a A) \cup \{B\}) = K *_a (A \wedge B)$.

We can use previous postulates to say more than AGM says in this case, namely, that

$$(K *_a A) *_a B = \mathrm{Cn}(K *_a A \cup \{B\}).$$

(1.2) B is inconsistent with $K_1 = K *_a A$, but B itself is consistent.

In this case, $\mathrm{Cn}(K *_a A \cup \{B\})$ is the set of all wffs.

- (K^*7) holds because the right hand side of the inclusion is the set of all wffs and any other set of formulae is included in this set.
- (K^*8) holds vacuously, since the antecedent of the implication is false.

(1.3) B is itself inconsistent.

- (K^*7) requires that $K *_a (A \wedge B) \subset \mathrm{Cn}((K *_a A) \cup \{B\})$ and
- (K^*8) holds vacuously.

The postulates say nothing new in this case, since the sets on either side of the inclusion in (K^*7) are equal to the set of all wffs of the language and (K^*8) is not applicable.

Case 2: A is not consistent with K, but A is itself consistent.

In this case, K_1 can be any consistent theory (by previous postulates), such that $A \in K_1$.

(2.1) B is consistent with K_1.

(2.2) B is inconsistent with K_1, but B itself is consistent.

(2.3) B is itself inconsistent.

These three cases follow, respectively, the same reasoning of cases (1.1), (1.2) and (1.3) above.

Case 3: A is itself inconsistent.

In this case, $K *_a A$ is the set of all wffs of the language. Whether or not B is consistent is irrelevant in the postulates in this case. $\mathrm{Cn}(K *_a A \cup \{B\})$ is equal to the set of all wffs and as for case (1.2) above

- (K*7) holds because any set of wff is included in $\mathrm{Cn}(K *_a A \cup \{B\})$.
- (K*8) holds vacuously, since the antecedent of the implication is false.

Summary of (K*7)–(K*8)

Postulates (K*7)–(K*8) do not tell us anything new (beyond what we can deduce from earlier postulates), except in the case where B is consistent with $K *_a A$ (case 1.1), when (K*7) and (K*8) together are equivalent to the postulate below:

($\mathrm{K}^*_{7,8}$) $\quad \mathrm{Cn}((K *_a A) \cup \{B\}) = K *_a (A \wedge B)$, when B is consistent with $K *_a A$

Therefore, for non-classical logics, we are committed only to ($\mathrm{K}^*_{7,8}$). Other cases involving inconsistency can have properties dictated by the local logic requirements.

(K*7) and (K*8) are the most interesting and controversial postulates. They capture in general terms the requirement that revisions are performed with a *minimal change* to the previous belief set. In order to understand them, recall that in a revision of K by A, one is interested in keeping as much as possible of the *informational content* of K and yet accept A. In semantical terms, this can be seen as looking for the models[4] of A that are somehow most *similar* to the models of the previous belief state K. The postulates do not constrain the operation well enough to give a precise meaning to the term similar, and this is how it should be, since they represent only general principles.

(K*7) says that if an interpretation I is among the models of A which are most similar to the models of K and it happens that I is also among the models of B, then I should also be among the models of $A \wedge B$ which are most similar to models of K.

[4] We consider models of a formula A, interpretations (or valuations) of the language which make A *true*.

Similarly, to understand the intuitive meaning of (K*8) consider the following situation: suppose that $(K*_a A)\wedge B$ is satisfiable. It follows that some models of A which are closest to models of K are also models of B. These models are obviously in $\mathrm{mod}(A \wedge B)$, since by (K*1), $\mathrm{mod}(K*_a A) \subseteq \mathrm{mod}(A)$. Now, every model in $\mathrm{mod}(A \wedge B)$ which is closest to models of K, must also be a model of $(K*_a A) \wedge B$.

This situation is depicted in Figure 1, where interpretations are represented around K according to their degree of similarity. The closer to $\mathrm{mod}(K)$ the more similar to K an interpretation is (the exact nature of this similarity notion is irrelevant to the understanding of the postulates).

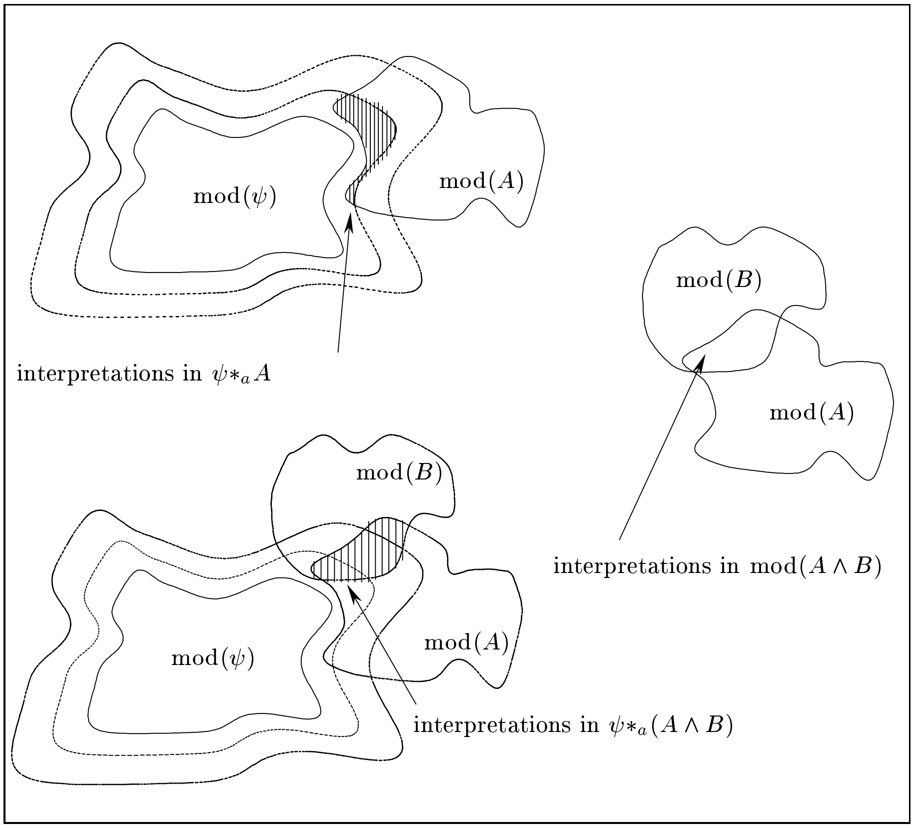

Figure 1: Illustrating postulate (K*8).

3 Revising in the modal logic K

We consider the case of the propositional modal logic K. The first thing we need to do is to provide the translation method from formulae and theories of K into formulae and theories of classical logic. This is done via the translation scheme described as follows.

Translation of modal K into classical logic

We need a binary predicate R in classical logic to represent the accessibility relation and unary predicates $P_1, P_2, P_3, \ldots$, for each propositional symbol p_i of K. We will use the subscript k whenever we wish to emphasize that we mean an operation (relation) in k and differentiate it from its classical logic counterpart (which will not be subscripted).

The idea is to encode the information of satisfiability of modal formulae by worlds into the variable of each unary predicate. In general, for a given world w and formula β the translation method can be stated as follows, where $\beta^\tau(w)$ represents $w \Vdash_k \beta$.

$$\begin{array}{rcl}
p_i^\tau(w) & = & P_i(w) \\
(\neg\beta)^\tau(w) & = & \neg(\beta^\tau(w)) \\
(\beta \wedge \gamma)^\tau(w) & = & \beta^\tau(w) \wedge \gamma^\tau(w) \\
(\beta \to \gamma)^\tau(w) & = & \beta^\tau(w) \to \gamma^\tau(w) \\
(\Box\beta)^\tau(w) & = & \forall y(wRy \to \beta^\tau(y))
\end{array}$$

Finally, for a modal theory Δ, we define

$$\Delta^\tau(w) = \{\beta^\tau(w) \mid \beta \in \Delta\}.$$

We have, where w_0 is a completely new constant to Δ and β, and represents the actual world, that:

$\Delta \vdash_k \beta$ iff in every Kripke model with actual world w_0, we have $w_0 \Vdash_k \Delta$ implies $w_0 \Vdash_k \beta$ iff in classical logic we have that $\Delta^\tau(w_0) \vdash \beta^\tau(w_0)$.
(Correspondence)

$$\Delta \vdash_k \beta \text{ iff } T^\tau \cup \Delta^\tau(w_0) \vdash \beta^\tau(w_0) \tag{2}$$

The theory T^τ in the case of the logic K is empty (i.e., Truth)[5].

If Δ is finite we can let $\delta = \bigwedge \Delta$ and we have

$$\delta \vdash_k \beta \text{ iff } \vdash \forall x(\delta^\tau(x) \to \beta^\tau(x))$$

We can define a revision operator $*_k$ for K, as outlined before (we will omit the reference to the actual world w_0 in the rest of this section).

Definition 2 [Belief revision in K]

$$\Delta *_k \psi = \{\alpha \mid \Delta^\tau *_a (\psi^\tau \wedge T^\tau) \vdash \alpha^\tau\}$$

We can now speak more specifically of properties of $*_k$:

[5] The logic K imposes no properties on R. Had we been translating $S4$, we would have $T^\tau = \{\forall x(xRx) \wedge \forall x \forall y \forall z(xRy \wedge yRz \to xRz)\}$. Our notion also allows for non-normal logics, e.g., if w_0 is the actual world, we can allow reflexivity in w_0 by setting $T^\tau = \{w_0 R w_0\}$.

Properties of $*_k$:

1. $\Delta *_k \psi$ is closed under $\vdash$.

   This can be easily shown.

2. $\Delta *_k \psi \vdash_k \psi$.

   By (K*2), $\psi^\tau \wedge T^\tau \in \Delta^\tau *_a (\psi^\tau \wedge T^\tau)$. Since $\Delta^\tau *_a (\psi^\tau \wedge T^\tau)$ is closed under $\vdash$, $\Delta^\tau *_a (\psi^\tau \wedge T^\tau) \vdash \psi^\tau$ and hence $\psi \in \Delta *_k \psi$, by (K*1), $\Delta *_k \psi \vdash_k \psi$.

3. If ψ is (modally) consistent with Δ, then $\Delta *_k \psi = \mathrm{Cn}_k(\Delta \cup \{\psi\})$.

   We first show that if ψ is modally consistent with Δ, then $\Delta^\tau(k)$ is classically consistent with $\psi^\tau(k) \wedge T^\tau$. This holds because $\Delta \cup \{\psi\}$ has a Kripke model which will give rise to a classical model of the translation. Therefore, $\Theta = \Delta^\tau(k) *_a (\psi^\tau(k) \wedge T^\tau)$ is the classical provability closure of $\Delta^\tau(k) \cup \{\psi^\tau(k) \wedge T^\tau\}$.

   We now have to show that if $\alpha^\tau(k) \in \theta$ then $\Delta *_k \psi \vdash \alpha$.

   Lemma 1 *Let Δ be a closed theory. Let Δ^τ be its translation and let $\mathrm{Cn}(\Delta^\tau)$ be its T^τ-closure in classical logic. Let β be such that $\beta^\tau \in \mathrm{Cn}(\Delta^\tau)$. It follows that $\beta \in \Delta$.*

   Proof: If $\beta \notin \Delta$, then there exists a Kripke model of $\Delta \cup \{\neg\beta\}$. This gives a classical model of $\Delta^\tau \cup \{T^\tau\} \cup \{\neg\beta^\tau\}$, and so $\beta^\tau \notin \mathrm{Cn}(\Delta^\tau)$. □

   Lemma 2 *Let Δ^τ be a closed classical theory such that $\Delta^\tau \vdash T^\tau$ and let $\Delta = \{\beta \mid \beta^\tau \in \Delta^\tau\}$. Then if $\Delta \vdash \alpha$, then $\alpha^\tau \in \Delta^\tau$.*

   Proof: If $\alpha^\tau \notin \Delta^\tau$, there exists a model of $\Delta^\tau \cup \{\neg\beta^\tau\}$. This can be viewed as a kripke model of $\Delta \cup \{\neg\beta\}$. □

4. $\Delta *_k \psi$ is modally inconsistent, only if ψ is modally contradictory.

   If $\Delta *_k \psi$ is inconsistent, then so is $\Delta^\tau *_a (\psi^\tau \wedge T^\tau)$, since $\Delta^\tau *_a (\psi^\tau \wedge T^\tau)$ is closed under $\vdash$. By (K*5), $\psi^\tau \wedge T^\tau$ is contradictory. But by the correspondence, $\psi^\tau \wedge T^\tau \vdash \bot$ iff $\psi \vdash_k \bot$.

5. If $\psi \equiv_k \varphi$, then $\Delta *_k \psi \equiv \Delta *_k \varphi$.

   By correspondence ((2), page 9) and since $\psi \vdash_k \varphi$ and $\varphi \vdash_k \psi$, it follows that $T^\tau \cup \psi^\tau \equiv T^\tau \cup \varphi^\tau$. Therefore, by (K*6), $\Delta *_a (T^\tau \cup \psi^\tau) \equiv \Delta *_a (T^\tau \cup \varphi^\tau)$ and hence $\Delta *_k \psi \equiv \Delta *_k \varphi$.

6. $\Delta *_k (\psi \wedge \varphi) \subseteq \mathrm{Cn}_k((\Delta *_k \psi) \cup \{\varphi\})$.

   Suppose that $\Delta *_k (\psi \wedge \varphi) \vdash_k \alpha$, for some α. By the definition of $*_k$, $\Delta^\tau *_a (\psi^\tau \wedge \varphi^\tau \wedge T^\tau) \vdash \alpha^\tau$. By (K*7), it follows that $\mathrm{Cn}(\Delta^\tau *_a (\psi^\tau \wedge T^\tau) \cup \{\varphi^\tau\} \vdash \alpha^\tau$. Notice that for every $\gamma^\tau \in \Delta^\tau *_a (\psi^\tau \wedge T^\tau)$, there is a corresponding γ in $\Delta *_k \psi$ (by the definition of $*_k$) and similarly for φ^τ. By correspondence, $\mathrm{Cn}_k((\Delta *_k \psi) \cup \{\varphi\}) \vdash \alpha$.

7. If φ is modally consistent with $\Delta *_k \psi$, then $\mathrm{Cn}((\Delta *_k \psi) \cup \{\varphi\}) \subseteq \Delta *_k (\psi \wedge \varphi)$.

   If φ is modally consistent with $\Delta *_k \psi$, then φ^τ is modally consistent with $\Delta *_a (\psi^\tau \wedge T^\tau)$ and then by (K*8), $\mathrm{Cn}(\Delta *_a (\psi^\tau \wedge T^\tau) \cup \{\varphi^\tau\}) \subseteq \mathrm{Cn}(\Delta^\tau *_a ((\psi \wedge \varphi)^\tau \wedge T^\tau))$. But, $\{\alpha \mid \Delta^\tau *_a (\psi^\tau \wedge T^\tau) \vdash \alpha^\tau\} \cup \{\varphi\} \vdash_k \beta$ iff $\Delta^\tau *_a (\psi^\tau \wedge T^\tau) \cup \{\varphi^\tau\} \vdash \beta^\tau$.

What we have just proven is that $*_k$ verifies all eight conditions of an AGM operation. Item 3, is actually a proof for postulates (K*3) and (K*4).

The actual process of revision briefly discussed in the example above may be more complex than this. For instance, the object language might have a consistency notion other than that of classical logic or even none at all. In Section 5, we examine such a case, by considering Belnap's four-valued logic.

4 Translating Belnap's four-valued logic into classical logic

Standard familiar systems such as classical logic, modal logic, intuitionistic logic have in common the principle that contradicting information entails any arbitrary sentence. This principle, known as *ex falsum quod libet*, is, however, not always appropriate to describe real application deduction processes, where information are often deduced from quite possibly inconsistent databases. Alternative systems have been developed, examples of which include the logic of *first-degree entailment* (also known as system **E**) and the *relevant implication* system (or system **R**), in which deductions between formulae hold only when there is some "connection" between the formulae (e.g. the formulae share some sentential variable). In [5] Belnap provides a semantic characterization of first-degree entailment together with a sound and complete axiomatisation, emphasising its connection with the problem of "how a computer should think" [4].

We provide a translation of Belnap's semantics into a set of first-order logic formulae. Sets of Belnap formulae are translated into a conjunction of atomic predicates. An appropriate classical axiomatisation is defined, which captures the semantic behaviour of Belnap connectives, thus allowing Belnap's notion of entailment to be expressed in terms of classical entailment from the translated theories. This embedding into classical logic has two main advantages. The first one is to provide the basis for analysing belief revision operations for these types of logics, as we have done for modal propositional logic in Section 3. Secondly, theorem provers for four-valued logic can be developed by applying existing classical theorem provers on the classical logic translation of these logics.

In Section 4.1, we illustrate Belnap's semantics, showing some of the features of the deduction process that it formalises and its differences with respect to familiar classically-based deductive systems. In Section 4.2, we define our translation approach and the classical axiomatization of Belnap's four-valued semantics, providing some illustrative derivation examples. We prove the soundness and completeness results of the translation approach, showing that it preserves Belnap's deductive process.

Notation & Terminology

We introduce specific notation as and when necessary throughout the rest of the paper. However, the reader might like to bear the following in mind: propositional symbols will usually begin with a lower-case letter, whereas predicate symbols will often begin with an upper–case letter. Greek–letter meta–variables will be used to refer in general to wffs of the Belnap logic (i.e. "object logic"), whereas upper-case meta-variable letters will be used to denote wffs of first-order logic (i.e. "target logic"). Larger entities such as structures, sets, theories and languages will often be symbolised in caligraphic font, $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$.

4.1 Belnap's four-valued Logic

As mentioned in [4], Belnap describes its semantic characterisation of four-valued logics as an appropriate logic for expressing practical deductive processes. In database management or question-answer systems, collections of data are proned to include either explicit or hidden inconsistencies. This is due for instance to the fact that information may come from different contradicting sources. The use of a classical deductive process would not be appropriate in this case – since any arbitrary information is classically derivable from an inconsistent collection of data. Explicit inconsistencies may come from different sources equally reliable, whereas hidden inconsistencies are identified only by means of deductive reasoning. The motivation for Belnap's approach is to provide a logic less sensitive to inconsistencies.

Syntax Let $\mathcal{L}_B$ be the Belnap propositional language composed of a countable set of propositional letters $\{p, q, r, \ldots\}$ and the connectives $\neg$, $\wedge$ and $\vee$. The set of wffs is given by the standard construction of formulae. For the finite case, a Belnap theory can be seen as a single formula given by the conjunction of a given finite set of wffs. The formula $\neg p \wedge (\neg q \vee r) \wedge \neg r$ is an example of a finite Belnap theory. Because of the soundness and completeness results of Section 4.2, it would not be difficult to extend this logic to deal with infinite theories, and we assume that such an extension exists in some of the proofs done in Section 5, where the AGM postulates are analysed.

In a proof theoretical terms, Belnap's four-valued logic is characterised by a finite axiomatization. Given two Belnap wffs α and β, the expression $\alpha \to \beta$ denotes that α *entails* β. In this sense, the symbol $\to$ can be seen as a derivability relation between formulae, or equally between theories and a formula. The expression $\alpha \leftrightarrow \beta$ denotes that β can be derived from α ($\alpha \to \beta$) and vice-versa, or, semantically, that α and β are equivalent. The axiomatization given below is known to be sound and complete with respect to the semantics of the logic presented later.

Definition 3 [Axiomatization] Let $\alpha_1, \ldots, \alpha_n$, $\beta_1, \ldots, \beta_m$ and γ be Belnap wffs. A proof theory for Belnap four-valued logic, denoted with Ax_B, is the following set of expressions:

1. $\neg\neg\alpha \leftrightarrow \alpha$.
2. $\neg(\alpha \wedge \beta) \leftrightarrow \neg\alpha \vee \neg\beta$.
3. $\neg(\alpha \vee \beta) \leftrightarrow \neg\alpha \wedge \neg\beta$.
4. $\alpha \vee \beta \leftrightarrow \beta \vee \alpha$.
5. $\alpha \vee (\beta \vee \gamma) \leftrightarrow (\alpha \vee \beta) \vee \gamma$.
6. $\alpha \vee (\beta \wedge \gamma) \leftrightarrow (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$.
7. $\alpha \wedge \beta \leftrightarrow \beta \wedge \alpha$.
8. $\alpha \wedge (\beta \wedge \gamma) \leftrightarrow (\alpha \wedge \beta) \wedge \gamma$.
9. $\alpha \wedge (\beta \vee \gamma) \leftrightarrow (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$.
10. $\alpha_1 \wedge \ldots \wedge \alpha_n \rightarrow \beta_1 \vee \ldots \vee \beta_m$ provided that $\beta_j = \alpha_i$ for some i and j.
11. $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$ then $\alpha \rightarrow \gamma$.
12. $\alpha \leftrightarrow \beta$ and $\beta \leftrightarrow \gamma$ then $\alpha \leftrightarrow \gamma$.
13. $\alpha \rightarrow \beta$ if and only if $\neg\beta \rightarrow \neg\alpha$.
14. $(\alpha \vee \beta) \rightarrow \gamma$ if and only if $\alpha \rightarrow \gamma$ and $\beta \rightarrow \gamma$.
15. $\alpha \rightarrow \beta$ if and only if $\beta \leftrightarrow (\alpha \vee \beta)$
16. $\alpha \rightarrow \beta$ if and only if $\alpha \leftrightarrow (\alpha \wedge \beta)$
17. $\alpha \rightarrow (\beta \wedge \gamma)$ if and only if $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$. $\square$

The first nine expressions correspond to standard classical properties of negation, disjunction and conjunction (e.g., commutativity, associativity, De Morgan laws). We will sometimes refer to them as the Belnap axioms. Expressions 10, 11 and 13 capture respectively the reflexivity, transitivity and contrapositive properties of the derivability relation $\rightarrow$, whereas expressions 14-17 correspond to standard classical rules for introduction and elimination of $\vee$ and $\wedge$ respectively. We will sometimes refer to expressions 11-17 as the Belnap rules. Any Belnap expression of the form $\psi \rightarrow \varphi$ can be either an instantiation of one of the axioms 1-10 in definition 3, or obtained using some of the Belnap rules 11-17, together with some axiom instantiations. For any given expression $\psi \rightarrow \varphi$, we therefore define the notion of *length* as the "least number" of Belnap rule applications needed to show $\psi \rightarrow \varphi$.

The similarity between the above rules and classical rules shows that four-valued logics are indeed very close to standard classical logic. The basic classical rule, which is missing in Belnap logic and which makes this logic *paraconsistent* is the rule $(\alpha \wedge \neg\alpha) \rightarrow \beta$, often referred to as *ex falsum quod libel.* This rule allows within a classical framework to derive any arbitrary information from inconsistent assumptions. Belnap logic does not allow so.

Semantics The semantics underlying Belnap's logic is four-valued. Let **4** be the set {T, F, Both, None}. The elements of this set are the four different truth-values which an atomic sentence can have within a given "state of information". The intuitive meaning of these values is given as follows:

1. p is stated to be true only (T)
2. p is stated to be false only (F)
3. p is stated to be both true and false, for instance, by different sources, or in different points of time (Both), and
4. p's status is unknown. That is, neither true, nor false (None).

The four values form a lattice, called the *approximation lattice* and denoted by **A4** where the ordering relation $\sqsubseteq$ goes "uphill" and respects the monotonicity property, in the sense that information about the truth-value of a formula "grows" from None to Both. **A4** can be seen in Figure 2.

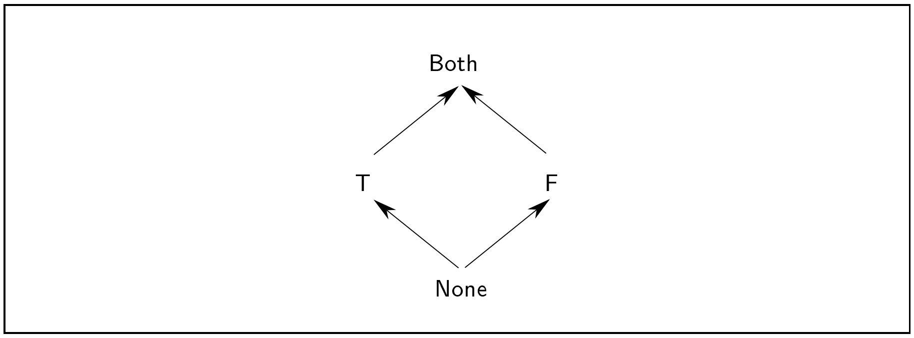

Figure 2: The approximation lattice **A4**.

The truth values of complex formulae are defined based on **A4** and result in the truth tables shown in Figure 3.

The truth tables constitute a lattice, called *logical lattice* and denoted by **L4** (Figure 4). In **L4**, logical conjunction is identified with the meet operation and logical disjunction with the join operation.

The notion of a interpretation of formulae is expressed in Belnap's logic in terms of *set-ups*. A set-up s is a mapping of the atomic formulae into **4**. Using the truth tables given in Figure 3, each set-up can be extended to a mapping of *all formulae* into **4**, in the standard inductive way. We call this extended set-up a **4**-*valuation* and denote it with v. Thus, for any given Belnap formula α and set-up s, the valuation $v(\alpha)$ is always well-defined. This makes Belnap's semantic somewhat different from the classical semantics, because the notion of model, that is, an interpretation that makes a formulae true is non-existent.

The notion of semantic entailment is then expressed in terms of a partial ordering $\preceq$ associated with the logical lattice **L4**. We will denote the semantic entailment relation with $\Rightarrow$ to distinguish it from the proof theoretic notion of entailment $\rightarrow$. The two notions are equivalent, as given by the correspondence 3, and the symbols $\rightarrow$ and $\Rightarrow$ will be often used interchangeably.

$\wedge$	None	F	T	Both
None	None	F	None	F
F	F	F	F	F
T	None	F	T	Both
Both	F	F	Both	Both

Truth-table for the connective $\wedge$

$\vee$	None	F	T	Both
None	None	None	T	T
F	None	F	T	Both
T	T	T	T	T
Both	T	Both	T	Both

Truth-table for the connective $\vee$

	None	F	T	Both
$\neg$	None	T	F	Both

Truth-table for the connective $\neg$

Figure 3: Truth-tables for Belnap's connectives.

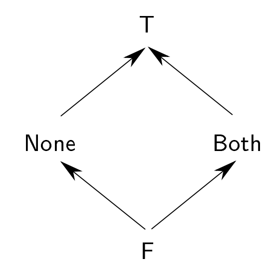

Figure 4: The logical lattice **L4**.

Definition 4 Let α and β be two Belnap formulae. We say that α *entails* β, written $\alpha \Rightarrow \beta$, if for all **4**-valuations v, $v(\alpha) \preceq v(\beta)$, where $\preceq$ is the partial ordering associated with the lattice **L4**. Analogously, a *non empty* finite set of formulae Γ *entails* α, if the conjunction of all formulae in Γ entails α.

$\square$

(Correspondence)

$$\alpha \to \beta \text{ iff } \alpha \Rightarrow \beta \tag{3}$$

Notice that if we restrict our attention to valuations into $\{\mathsf{F}, \mathsf{T}\}$ only, we get the familiar classic logic notions. We need a parallel to the notion of consistency, this would be *acceptability*. We say a Belnap theory Δ is *acceptable*, if for any wff α, Δ does not prove both α and $\neg\alpha$. This is *our* definition and it seems reasonable to us. A consistent/acceptable theory can tell us that a given α is true (T), false (F) or don't know (None), but if it says that it is both true and false (Both), then something in the knowledge about α went wrong.

Definition 5 [Acceptability of Belnap theories] A theory Δ is *acceptable*, if

$$\{\gamma \mid \Delta \rightarrow \gamma \text{ and } \Delta \rightarrow \neg\gamma\} = \varnothing$$

$\square$

Note that we do not have that if Δ is not acceptable, then it can prove everything. We now introduce some terminology which will be used throughout this paper.

Definition 6 Let α be a Belnap formula and let v be a **4**-valuation. We say that α is

- *at least true* under v if $v(\alpha) = \mathsf{T}$ or $v(\alpha) = \mathsf{Both}$.
- *at least false* under v if $v(\alpha) = \mathsf{F}$ or $v(\alpha) = \mathsf{Both}$.
- *not true* under v if $v(\alpha) = \mathsf{F}$ or $v(\alpha) = \mathsf{None}$.
- *not false* under v if $v(\alpha) = \mathsf{T}$ or $v(\alpha) = \mathsf{None}$. $\square$

Using the above terminology, the notion of semantic entailment between a theory and a formula given in Definition 4 can be equivalently expressed as follows.

Definition 7 Let Γ be a set of Belnap formulae and α a Belnap formula. Γ entails α if and only if for every **4**-valuation v,

i) if all the formulae in Γ are at least true under v, then α is at least true under v;

ii) if all the formulae in Γ are not false under v, then α is not false under v. $\square$

This definition will play an important role in the soundness and completeness proofs of the first-order Belnap translation with respect to Belnap semantics.

4.2 The translation into classical logic

In this section, we describe a translation approach of Belnap logic into first-order logic and show that it is sound and complete with respect to Belnap's semantic notion of entailment. Let $\mathcal{L}$ be a two sorted first-order language composed of the sort $\mathcal{F}$, called *B-formulae*, and the sort $\mathcal{V}$ called *truth values*.

The set of constants of the sort $\mathcal{F}$ is the set of propositional letters in Belnap's logic, whereas terms of $\mathcal{F}$ are constructed using three main functions $\neg$, $\wedge$, and $\vee$ which correspond to the Belnap connectives. The set of ground terms of $\mathcal{F}$ is therefore equivalent to the set of Belnap wffs. The sort $\mathcal{V}$ is instead composed of two constant symbols $\{\mathsf{tt}, \mathsf{ff}\}$, the basic constants from which Belnap's four-valued semantics can be constructed. $\mathcal{L}$ also contains the two-sorted binary predicate *holds*. *holds* takes as first arguments, $\mathcal{F}$ terms, and as second arguments $\mathcal{V}$ terms. $\mathcal{F}$ variables will be denoted with x, y, z, ... First-order formulae are constructed in the usual way.

Ground atomic formulae can be of two types $holds(\varphi, \mathsf{tt})$ and $holds(\psi, \mathsf{ff})$ for any Belnap wffs φ and ψ. Atomic formulae of the first type mean that "$\mathsf{tt} \in v(\varphi)$", for some **4**-valuation v, which is equivalent to say that for some **4**-valuation, φ is at least true. Atomic formulae of the second type state instead that "$\mathsf{ff} \in v(\psi)$", for some **4**-valuation v, which is equivalent to say that for some **4**-valuation, ψ is at least false. With these two types of atomic formulae it is possible to express Belnap's full four-valued semantics. In order to simplify the proof, we extend the sort $\mathcal{V}$ with four constant symbols T , F , None and Both, as follows:

Definition 8 Let α be a Belnap formula. The four truth values that α can assume in Belnap semantics is expressed in the first-order translation by the following additional types of atomic formulae:

$$
\begin{aligned}
holds(\alpha, \mathsf{T}) &\stackrel{\text{def}}{=} holds(\alpha, \mathsf{tt}) \wedge \neg holds(\alpha, \mathsf{ff}) \\
holds(\alpha, \mathsf{F}) &\stackrel{\text{def}}{=} \neg holds(\alpha, \mathsf{tt}) \wedge holds(\alpha, \mathsf{ff}) \\
holds(\alpha, \mathsf{None}) &\stackrel{\text{def}}{=} \neg holds(\alpha, \mathsf{tt}) \wedge \neg holds(\alpha, \mathsf{ff}) \\
holds(\alpha, \mathsf{Both}) &\stackrel{\text{def}}{=} holds(\alpha, \mathsf{tt}) \wedge holds(\alpha, \mathsf{ff})
\end{aligned}
$$

□

The atomic formulae on the left-hand side express that under a **4**-valuation v, $v(\alpha) = \mathsf{T}$, $v(\alpha) = \mathsf{F}$, $v(\alpha) = \mathsf{None}$ and $v(\alpha) = \mathsf{Both}$ respectively. However, these additional four types of atomic formulae are in reality a short-hand for first-order formulae constructed from the basic language $\mathcal{L}$. We will therefore use throughout the paper only the basic atomic formulae of $\mathcal{L}$.

The semantic behavior of Belnap connectives is fully captured by the following first-order axiomatisation.

Definition 9 Given the two languages $\mathcal{L}_B$ and $\mathcal{L}$, $\mathcal{A}_B$ is the first-order axiomatisation of Belnap four-valued semantics given by the following six axioms:

$$\forall x[holds(x, \mathsf{ff}) \leftrightarrow holds(\neg x, \mathsf{tt})] \tag{Ax 1}$$
$$\forall x[holds(x, \mathsf{tt}) \leftrightarrow holds(\neg x, \mathsf{ff})] \tag{Ax 2}$$
$$\forall x, y[holds(x \wedge y, \mathsf{tt}) \leftrightarrow (holds(x, \mathsf{tt}) \wedge holds(y, \mathsf{tt}))] \tag{Ax 3}$$
$$\forall x, y[holds(x \wedge y, \mathsf{ff}) \leftrightarrow (holds(x, \mathsf{ff}) \vee holds(y, \mathsf{ff}))] \tag{Ax 4}$$
$$\forall x, y[holds(x \vee y, \mathsf{tt}) \leftrightarrow (holds(x, \mathsf{tt}) \vee holds(y, \mathsf{tt}))] \tag{Ax 5}$$
$$\forall x, y[holds(x \vee y, \mathsf{ff}) \leftrightarrow (holds(x, \mathsf{ff}) \wedge holds(y, \mathsf{ff}))] \tag{Ax 6}$$

□

The *translation function* τ is a mapping from the set of Belnap wffs to the set of ground atomic first-order formulae of the form $holds(\varphi, \mathsf{tt})$. For a given Belnap formula φ, its first order translation, denoted with $\tau(\varphi)$ or simply φ^τ, is the first-order atomic formula $holds(\varphi, \mathsf{tt})$. The translation of a Belnap theory (i.e. finite sets of Belnap formulae) is therefore given by the translation of the conjunction of all Belnap formulae included in the theory. For instance, let

$\Gamma = \{\varphi_1, \ldots, \varphi_n\}$ be a Belnap theory, its translation $\tau(\Gamma)$, or Γ^τ, is the atomic first-order formula $holds(\varphi_1 \wedge \ldots \wedge \varphi_n, \mathsf{tt})$.

We are now going to prove that the above translation function together with the axiomatisation $\mathcal{A}_B$ is sound and complete with respect to the Belnap semantic notion of entailment.

Theorem 1 (Correspondence) *Let ψ and φ be two Belnap formulae.*

$$\begin{array}{lll} \psi \to \varphi & \text{iff} & \mathcal{A}_B, holds(\psi, \mathsf{tt}) \vdash holds(\varphi, \mathsf{tt}) \\ & \text{and} & \mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \vdash \neg holds(\varphi, \mathsf{ff}). \end{array}$$

The proof of the above theorem uses Lemmas 3 and 4. Lemma 3 expresses the completeness of the translation function and the first-order axiomatisation with respect to the notion of Belnap entailment.

The statement captures, in first-order terms, the notion of entailment, given in Definition 7, whenever ψ is of the form $\alpha_1 \wedge \ldots \wedge \alpha_n$, with $\{\alpha_1, \ldots, \alpha_n\}$ being a Belnap theory.

For the first conjunct of the statement, the assumption $holds(\psi, \mathsf{tt})$ is equivalent, by axiom (Ax 3), to $holds(\alpha_1, \mathsf{tt}) \wedge \ldots \wedge holds(\alpha_n, \mathsf{tt})$, which can be read as "all α_i, for each $1 \leq i \leq n$, are at least true". The consequence $holds(\varphi, \mathsf{tt})$ can also be read as φ is at least true. Analogously, for the second conjunct in the statement, the assumption $\neg holds(\psi, \mathsf{ff})$ is equivalent, by axiom (Ax 4), to $\neg holds(\alpha_1, \mathsf{ff}) \wedge \ldots \wedge \neg holds(\alpha_n, \mathsf{ff})$, where each $\neg holds(\alpha_i, \mathsf{ff})$ can be read as "α_i is not false". Lemma 4 expresses instead the soundness of the translation function and the first-order axiomatisation with respect to Belnap's notion of entailment.

Lemma 3 (Completeness) *Let ψ and φ be two Belnap formulae.*

$$\begin{array}{lll} \text{If } \psi \to \varphi \text{ then} & \mathcal{A}_B,\ holds(\psi, \mathsf{tt}) \ \vdash\ holds(\varphi, \mathsf{tt}) & \text{and} \\ & \mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \ \vdash \neg holds(\varphi, \mathsf{ff}). & \end{array}$$

Proof: The proof is by induction on the length n of the derivation $\psi \to \varphi$.
Base Case: $n = 0$. Then $\psi \to \varphi$ can only be an instantiation of one of the axioms 1-10 given in Definition 3. The proof is therefore by cases on each of these axioms. Only some of the cases are shown here. The remaining ones are proved following the same type of argument.
Case 1: $\psi \to \varphi$ is an instantiation of $\alpha_1 \wedge \ldots \wedge \alpha_h \to \beta_1 \vee \ldots \vee \beta_k$, for some h and k such that $\alpha_i = \beta_j$ for some i and j. We show in Figure 5 that $\mathcal{A}_B, holds(\alpha_1 \wedge \ldots \wedge \alpha_h, \mathsf{tt}) \vdash holds(\beta_1 \vee \ldots \vee \beta_k, \mathsf{tt})$ and in Figure 6 that $\mathcal{A}_B, \neg holds(\alpha_1 \wedge \ldots \wedge \alpha_h, \mathsf{ff}) \vdash \neg holds(\beta_1 \vee \ldots \vee \beta_k, \mathsf{ff})$.
Case 2: $\psi \to \varphi$ is an instantiation of $\alpha \vee (\beta \wedge \gamma) \to (\alpha \vee \beta) \wedge (\alpha \wedge \gamma)$. We show in Figure 8 that $\mathcal{A}_B, holds(\alpha \vee (\beta \wedge \gamma), \mathsf{tt}) \to holds((\alpha \vee \beta) \wedge (\alpha \wedge \gamma), \mathsf{tt})$, and in Figure 7 that $\mathcal{A}_B, \neg holds(\alpha \vee (\beta \wedge \gamma), \mathsf{ff}) \to \neg holds((\alpha \vee \beta) \wedge (\alpha \wedge \gamma), \mathsf{ff})$. Similar argument is applied in the case where $\psi \to \varphi$ is an instantiation of $(\alpha \vee \beta) \wedge (\alpha \wedge \gamma) \to \alpha \vee (\beta \wedge \gamma)$.
Inductive Step: We assume that there exists a first part of a derivation proving an expression of the form $\alpha \to \beta$ with $n - 1$ applications of Belnap rules; and that the n-th application of a Belnap rule gives us the expression $\psi \to \varphi$. We

$$
\begin{array}{ll}
\underline{\mathcal{A}_B, holds(\alpha_1 \wedge \ldots \wedge \alpha_h, \mathsf{tt})} & (\text{Ax } 3)\\
\underline{holds(\alpha_1, \mathsf{tt}) \wedge \ldots \wedge holds(\alpha_h, \mathsf{tt})} & (\mathcal{E}\wedge)\\
\underline{holds(\alpha_i, \mathsf{tt})} & (\text{equiv. rewriting})\\
\underline{holds(\beta_j, \mathsf{tt})} & (\mathcal{I}\vee)\\
\underline{holds(\beta_1, \mathsf{tt}) \vee \ldots \vee holds(\beta_k, \mathsf{tt})} & (\text{Ax } 5)\\
holds(\beta_1 \vee \ldots \vee \beta_k, \mathsf{tt}) &
\end{array}
$$

Figure 5: First-order proof of Belnap axiom 10.

$$
\begin{array}{ll}
\underline{\mathcal{A}_B, \neg holds(\alpha_1 \wedge \ldots \wedge \alpha_h, \mathsf{ff})} & \\
\quad \underline{holds(\beta_1 \vee \ldots \vee \beta_k, \mathsf{ff})} & (\text{assumption})\\
\quad \underline{holds(\beta_1, \mathsf{ff}) \wedge \ldots \wedge holds(\beta_k, \mathsf{ff})} & (\text{Ax } 6)\\
\quad \underline{holds(\beta_j, \mathsf{ff})} & (\mathcal{E}\wedge)\\
\quad \underline{holds(\alpha_i, \mathsf{ff})} & (\text{equiv.rewriting})\\
\quad \underline{holds(\alpha_1, \mathsf{ff}) \vee \ldots \vee holds(\alpha_h, \mathsf{ff})} & (\mathcal{I}\vee)\\
\quad \underline{holds(\alpha_1 \wedge \ldots \wedge \alpha_h, \mathsf{ff})} & (\text{Ax } 4)\\
\quad \underline{\perp} & (\mathcal{I}\neg)\\
\neg holds(\beta_1 \vee \ldots \vee \beta_k, \mathsf{ff}) &
\end{array}
$$

Figure 6: First-order proof of Belnap axiom 10.

reason by cases on each Belnap rule that could have been applied on this n-th step.
Case 1: We assume that last rule application is the "if-part" of Belnap rule 13 in Definition 3. Therefore, we have that there exists a proof of $\neg\varphi \to \neg\psi$, with $n-1$ rule applications. So by inductive hypothesis we can say that $\mathcal{A}_B, holds(\neg\varphi, \mathsf{tt}) \vdash holds(\neg\psi, \mathsf{tt})$ and that $\mathcal{A}_B, \neg holds(\neg\varphi, \mathsf{ff}) \vdash \neg holds(\neg\psi, \mathsf{ff})$. We want then to show that

$$\mathcal{A}_B, holds(\psi, \mathsf{tt}) \vdash holds(\varphi, \mathsf{tt})$$

and that

$$\mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \vdash \neg holds(\varphi, \mathsf{ff}).$$

From the inductive hypothesis $\mathcal{A}_B, \neg holds(\neg\varphi, \mathsf{ff}) \vdash \neg holds(\neg\psi, \mathsf{ff})$, we get,

$$
\begin{array}{ll}
\underline{\mathcal{A}_B, \neg holds(\alpha \vee (\beta \wedge \gamma), \mathsf{ff})} & \text{(Ax 6)} \\
\underline{\neg holds(\alpha, \mathsf{ff}) \vee \neg holds(\beta \wedge \gamma, \mathsf{ff})} & \text{(Ax 4)} \\
\underline{\neg holds(\alpha, \mathsf{ff}) \vee (\neg holds(\beta, \mathsf{ff}) \wedge \neg holds(\gamma, \mathsf{ff}))} & \text{(De Morgan Law)} \\
\underline{(\neg holds(\alpha, \mathsf{ff}) \vee \neg holds(\beta, \mathsf{ff})) \wedge (\neg holds(\alpha, \mathsf{ff}) \vee \neg holds(\gamma, \mathsf{ff}))} & \text{(Ax 6)} \\
\underline{\neg holds(\alpha \vee \beta, \mathsf{ff}) \wedge \neg holds(\alpha \vee \gamma, \mathsf{ff})} & \text{(Ax 4)} \\
\neg holds((\alpha \vee \beta) \wedge (\alpha \vee \gamma), \mathsf{ff}) &
\end{array}
$$

Figure 7: First-order proof of left-to-right part of Belnap axiom 6.

$$
\begin{array}{ll}
\underline{\mathcal{A}_B, holds(\alpha \vee (\beta \wedge \gamma), \mathsf{tt})} & \text{(Ax 5)} \\
\underline{holds(\alpha, \mathsf{tt}) \vee holds(\beta \wedge \gamma, \mathsf{tt})} & \text{(Ax 3)} \\
\underline{holds(\alpha, \mathsf{tt}) \vee (holds(\beta, \mathsf{tt}) \wedge holds(\gamma, \mathsf{tt}))} & \text{(De Morgan Law)} \\
\underline{(holds(\alpha, \mathsf{tt}) \vee holds(\beta, \mathsf{tt})) \wedge (holds(\alpha, \mathsf{tt}) \vee holds(\gamma, \mathsf{tt}))} & \text{(Ax 5)} \\
\underline{holds(\alpha \vee \beta, \mathsf{tt}) \wedge holds(\alpha \vee \gamma, \mathsf{tt})} & \text{(Ax 3)} \\
holds((\alpha \vee \beta) \wedge (\alpha \vee \gamma), \mathsf{tt}) &
\end{array}
$$

Figure 8: First-order proof of left-to-right part of Belnap axiom 6.

by contrapositive of classical logic, that

$$\mathcal{A}_B, holds(\neg\psi, \mathsf{ff}) \vdash holds(\neg\varphi, \mathsf{ff})$$

Hence, using Belnap axiom 2, we get $\mathcal{A}_B, holds(\psi, \mathsf{tt}) \vdash holds(\varphi, \mathsf{tt})$. To show that $\mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \vdash \neg holds(\varphi, \mathsf{ff})$ we consider the second part of the inductive hypothesis. $\mathcal{A}_B, holds(\neg\varphi, \mathsf{tt}) \vdash holds(\neg\psi, \mathsf{tt})$ gives, by contrapositive of classical logic that

$$\mathcal{A}_B, \neg holds(\neg\psi, \mathsf{tt}) \vdash \neg holds(\neg\varphi, \mathsf{tt}).$$

Hence, by Belnap axiom 1, $\mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \vdash \neg holds(\varphi, \mathsf{ff})$. The case for the "only if-part" of Belnap rule 13 follows the same argument.

Case 2: We assume that last rule application is the "if-part" of Belnap rule 15 in Definition 3. Therefore, we have there exists a proof of $\alpha \rightarrow \beta$ with $n-1$ rule applications, where ψ is equal to α and φ is equal to $\alpha \vee \beta$. So by inductive hypothesis, $\mathcal{A}_B, holds(\alpha, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt})$ and that $\mathcal{A}_B, \neg holds(\alpha, \mathsf{ff}) \vdash \neg holds(\beta, \mathsf{ff})$. We want to show that

1. $\mathcal{A}_B, holds(\beta, \mathsf{tt}) \vdash holds(\alpha \vee \beta, \mathsf{tt})$ and
$\mathcal{A}_B, holds(\alpha \vee \beta, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt})$
2. $\mathcal{A}_B, \neg holds(\beta, \mathsf{ff}) \vdash \neg holds(\alpha \vee \beta, \mathsf{ff})$ and
$\mathcal{A}_B, \neg holds(\alpha \vee \beta, \mathsf{ff}) \vdash \neg holds(\beta, \mathsf{ff})$.

The first part of (1) is quite straightforward. We show the second part. Assume $\mathcal{A}_B, holds(\alpha \vee \beta, \mathsf{tt})$. By axiom (Ax 5) and reflexivity of classical logic, $\mathcal{A}_B, holds(\alpha \vee \beta, \mathsf{tt}) \vdash \mathcal{A}_B, holds(\alpha, \mathsf{tt}) \vee holds(\beta, \mathsf{tt})$. By inductive hypothesis, $\mathcal{A}_B, holds(\alpha, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt})$ and by reflexivity of classical logic

$$\mathcal{A}_B, holds(\beta, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt}).$$

Therefore, using classical $\vee$-introduction rule,

$$\mathcal{A}_B, holds(\alpha, \mathsf{tt}) \vee holds(\beta, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt}).$$

Hence, $\mathcal{A}_B, holds(\alpha \vee \beta, \mathsf{tt}) \vdash holds(\beta, \mathsf{tt})$. The proof for (2) follows the same argument.
All the other cases can be easily proved using appropriate properties and rules of classical logic and, if necessary, the Belnap axioms. $\square$

Lemma 4 (Soundness) *Let ψ and φ be two Belnap formulae.*
If $\mathcal{A}_B, holds(\psi, \mathsf{tt}) \vdash holds(\varphi, \mathsf{tt})$ and $\mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \vdash \neg holds(\varphi, \mathsf{ff})$, then $\psi \rightarrow \varphi$.

Some additional propositions and definitions need to be given before proving the above lemma. The soundness of the classical translation is based on the idea that for any given Belnap **4**-valuation it is always possible to construct a classical interpretation I which satisfies the classical axioms $\mathcal{A}_B$ and which preserves Belnap's semantic entailment. We show first how this classical interpretation can be constructed and its properties.

Definition 10 Let v be a Belnap **4**-valuation from the set of Belnap wffs to the power set $\wp(\{\mathsf{tt}, \mathsf{ff}\})$. A classical interpretation *associated* with v, and denoted with $\mathcal{I}_v$, is a function defined as follows

- $\mathcal{I}_v(\mathsf{tt}) = \mathsf{tt}$ and $\mathcal{I}_v(\mathsf{ff}) = \mathsf{ff}$.

Also, for each ground term α of sort $\mathcal{F}$:

- $\mathcal{I}_v(\alpha) = \alpha$, for each ground term α of sort $\mathcal{F}$.
- $\mathcal{I}_v(holds) = \{\langle \alpha, \mathsf{tt} \rangle \mid \mathsf{tt} \in v(\alpha)\} \cup \{\langle \alpha, \mathsf{ff} \rangle \mid \mathsf{ff} \in v(\alpha))\}$ $\square$

It is easy to show, by definition of $\mathcal{I}_v$, that the following properties hold for any Belnap formula α and **4**-valuation v.

- $v(\alpha) = \mathsf{T}$ if and only if $\mathcal{I}_v \models holds(\alpha, \mathsf{tt}) \wedge \neg holds(\alpha, \mathsf{ff})$
- $v(\alpha) = \mathsf{F}$ if and only if $\mathcal{I}_v \models holds(\alpha, \mathsf{ff}) \wedge \neg holds(\alpha, \mathsf{tt})$

- $v(\alpha) = \mathsf{Both}$ if and only if $\mathcal{I}_v \models holds(\alpha, \mathsf{tt}) \wedge holds(\alpha, \mathsf{ff})$
- $v(\alpha) = \mathsf{None}$ if and only if $\mathcal{I}_v \models \neg holds(\alpha, \mathsf{tt}) \wedge \neg holds(\alpha, \mathsf{ff})$

The following proposition shows that a classical interpretation $\mathcal{I}_v$ associated to a given **4**-valuation v is a model of the first-order axioms $\mathcal{A}_B$.

Proposition 1 *Let v be a **4**-valuation and let $\mathcal{I}_v$ be its associated classical interpretation. Then $\mathcal{I}_v$ is a model of the classical axiomatisation $\mathcal{A}_B$.*

Proof: The proof is by cases of each axiom of $\mathcal{A}_B$.

Case 1: (Ax 1). We want to show that $\mathcal{I}_v \models \forall x[holds(x, \mathsf{ff}) \leftrightarrow holds(\neg x, \mathsf{tt})]$. We reason by contradiction. We assume that, for some x, $\mathcal{I}_v \models holds(x, \mathsf{ff})$ and $\mathcal{I}_v \not\models holds(\neg x, \mathsf{tt})$. By definition of $\mathcal{I}_v$, $\mathsf{ff} \in v(x)$, which implies by the $\neg$ truth table that $\mathsf{tt} \in v(\neg x)$. Hence $\mathcal{I}_v \models holds(\neg x, \mathsf{tt})$ which contradicts the hypothesis. Similarly for the other case, i.e. $\mathcal{I}_v \not\models holds(x, \mathsf{ff})$ and $\mathcal{I}_v \models holds(\neg x, \mathsf{tt})$.

Case 3: (Ax 3). We want to show that $\mathcal{I}_v \models \forall x, y[holds(x \wedge y, \mathsf{tt}) \leftrightarrow (holds(x, \mathsf{tt}) \wedge holds(y, \mathsf{tt}))]$. We reason by contradiction. Assume that, for some x, $\mathcal{I}_v \models holds(x \wedge y, \mathsf{tt})$, and $\mathcal{I}_v \not\models holds(x, \mathsf{tt})$ or $\mathcal{I}_v \not\models holds(y, \mathsf{tt})$. By definition of $\mathcal{I}_v$, $\mathsf{tt} \in v(x \wedge y)$, which implies by the $\wedge$ truth table that $\mathsf{tt} \in v(x)$ and $\mathsf{tt} \in v(y)$. Therefore, $\mathcal{I}_v \models holds(x, \mathsf{tt})$ and $\mathcal{I}_v \models holds(y, \mathsf{tt})$, which is in contradiction with the initial hypothesis. The second case, i.e. assume that, for some x, $\mathcal{I}_v \not\models holds(x \wedge y, \mathsf{tt})$, and $\mathcal{I}_v \models holds(x, \mathsf{tt})$ and $\mathcal{I}_v \models holds(y, \mathsf{tt})$, can be proved following the same argument.

Case 5: (Ax 5). We want to show that $\mathcal{I}_v \models \forall x, y[holds(x \vee y, \mathsf{tt}) \leftrightarrow (holds(x, \mathsf{tt}) \vee holds(y, \mathsf{tt}))]$. We reason by contradiction. Assume that, for some x, $\mathcal{I}_v \models holds(x \vee y, \mathsf{tt})$, and $\mathcal{I}_v \not\models holds(x, \mathsf{tt})$ and $\mathcal{I}_v \not\models holds(y, \mathsf{tt})$. By definition of $\mathcal{I}_v$, $\mathsf{tt} \in v(x \vee y)$, which implies by the $\vee$ truth table that $\mathsf{tt} \in v(x)$ or $\mathsf{tt} \in v(y)$. Therefore, $\mathcal{I}_v \models holds(x, \mathsf{tt})$ or $\mathcal{I}_v \models holds(y, \mathsf{tt})$, which is in contradiction with the initial hypothesis. The second case, i.e. assume that, for some x, $\mathcal{I}_v \not\models holds(x \vee y, \mathsf{tt})$, and $\mathcal{I}_v \models holds(x, \mathsf{tt})$ or $\mathcal{I}_v \models holds(y, \mathsf{tt})$, can be proved following the same argument.

Axioms 2,4 and 6 are proved in an analogous way of the proofs of Axioms, 1,3 and 5, respectively. $\square$

Proof of Lemma 4. We prove the contrapositive statement. We assume that $\psi \not\rightarrow \varphi$ and we want to show that either $\mathcal{A}_B, holds(\psi, \mathsf{tt}) \not\vdash holds(\varphi, \mathsf{tt})$ or $\mathcal{A}_B, \neg holds(\psi, \mathsf{ff}) \not\vdash \neg holds(\varphi, \mathsf{ff})$. The hypothesis $\psi \not\rightarrow \varphi$ implies different cases or truth values for ψ and φ according to the ordering relation $\preceq$ over the logical lattice **L4**. We consider these cases individually. $\psi \not\rightarrow \varphi$ implies that for some **4**-valuation v, $v(\psi) \not\preceq v(\varphi)$.

Case 1: $v(\psi) = \mathsf{T}$ and $v(\varphi) = \mathsf{Both}$. From v, we can construct the associated classical interpretation $\mathcal{I}_v$. By definition, $\mathcal{I}_v \models holds(\psi, \mathsf{tt})$ and $\mathcal{I}_v \models \neg holds(\psi, \mathsf{ff})$. But $\mathcal{I}_v \not\models \neg holds(\varphi, \mathsf{ff})$.

Case 2: $v(\psi) = \mathsf{T}$ and $v(\varphi) = \mathsf{None}$. Then $\mathsf{tt} \notin v(\varphi)$. From v, we can construct the associated classical interpretation $\mathcal{I}_v$. By definition, $\mathcal{I}_v \models \mathit{holds}(\psi, \mathsf{tt})$ and $\mathcal{I}_v \models \neg\mathit{holds}(\psi, \mathsf{ff})$. But $\mathcal{I}_v \not\models \mathit{holds}(\varphi, \mathsf{tt})$.

Case 3: $v(\psi) = \mathsf{T}$ and $v(\varphi) = \mathsf{F}$. Then $\mathsf{tt} \notin v(\varphi)$. From v, we can construct the associated classical interpretation $\mathcal{I}_v$. By definition, $\mathcal{I}_v \models \mathit{holds}(\psi, \mathsf{tt})$ and $\mathcal{I}_v \models \neg\mathit{holds}(\psi, \mathsf{ff})$. But $\mathcal{I}_v \not\models \mathit{holds}(\varphi, \mathsf{tt})$.

Case 4: $v(\psi) = \mathsf{None}$ and $v(\varphi) = \mathsf{F}$. Then $\mathsf{tt} \notin v(\psi)$, $\mathsf{ff} \notin v(\psi)$ and $\mathsf{tt} \notin v(\varphi)$. From v, we can construct the associated classical interpretation $\mathcal{I}_v$. By definition, $\mathcal{I}_v \models \neg\mathit{holds}(\psi, \mathsf{ff})$, but but $\mathcal{I}_v \not\models \neg\mathit{holds}(\varphi, \mathsf{ff})$.

Case 5: $v(\psi) = \mathsf{Both}$ and $v(\varphi) = \mathsf{F}$. Then $\mathsf{tt} \in v(\psi)$, $\mathsf{ff} \in v(\psi)$ and $\mathsf{tt} \notin v(\varphi)$. From v, we can construct the associated classical interpretation $\mathcal{I}_v$. By definition, $\mathcal{I}_v \models \mathit{holds}(\psi, \mathsf{tt})$, but but $\mathcal{I}_v \not\models \mathit{holds}(\varphi, \mathsf{tt})$. □

Proof of Theorem 1. The "if-part" is given by Lemma 4 whereas the "only-if" part is given by Lemma 3. □

5 Revising Belnap's four-valued logic

The main idea of doing Belief Revision by translation is that a revision operator for classical logic can be used for other logics as well, as long as a sound and complete translation of the target language into classical logic is given. Given a theory in the target language, the revision is performed on its translation together with the axioms which describe the behaviour of the logic in classical logic.

In this section we illustrate this idea by considering in depth the steps necessary to implement it. We take as an example the Belnap's logic discussed in the previous sections.

The notion of consistency play therefore a very important role in the process. In fact, one of the postulates ensures that the result of the revision process is always consistent provided that the new information is itself inconsistent.

In order to perform belief revision in one logic through its translation to classical logic, it is necessary to ensure that the notion of inconsistency in the object language is translated into inconsistency in classical logic. More precisely, it is not even necessary to be limited to the notion of inconsistency in the object language. As a matter of fact, some logics do not even have such a notion, e.g., Belnap's logic seen in the previous section. In such cases, one must define a notion of acceptability of object theories via classical logic axioms.

Take for instance Belnap's theories. We have defined a Belnap theory Δ *acceptable*, if for no α, it is the case that $\Delta \to \alpha$ and $\Delta \to \alpha$ (Definition 5).

This notion of acceptability is then be translated into classical logic inconsistency to ensure that revisions of Belnap's theories performed in classical logic will correspond to *acceptable* Belnap theories. The notion of acceptability seen above is translated into classical logic in the following way:

Definition 11 Acceptability in the translation to classical logic is defined as

$$Acc = \forall\alpha[(holds(\alpha, \mathsf{tt}) \wedge holds(\alpha, \mathsf{ff})) \rightarrow \bot]$$

As we expected, the co-existence of contradictory (Belnap) information will not be tolerated by the revision process, and only one of tv or fv (or none) will be accepted. Intuitively, this filters out Belnap formulae with truth-value Both.

Let ψ and φ be two Belnap formulae. The revision of ψ by φ, in symbols $\psi *_b \varphi$ is defined in terms of a First-order AGM compliant revision operator $*_a$ in the following way (we use the previously mentioned notion of *acceptability* Acc). In order to lighten the notation, we use the following convention:

$$\begin{array}{lcllcl}
holds(\alpha, \mathsf{tt}) & = & \alpha^+ & \neg holds(\alpha, \mathsf{tt}) & = & \neg\alpha^+ \\
holds(\alpha, \mathsf{ff}) & = & \alpha^- & \neg holds(\alpha, \mathsf{ff}) & = & \neg\alpha^- \\
holds(\neg\alpha, \mathsf{tt}) & = & \overline{\alpha}^+ & \neg holds(\neg\alpha, \mathsf{tt}) & = & \neg\overline{\alpha}^+ \\
holds(\neg\alpha, \mathsf{ff}) & = & \overline{\alpha}^- & \neg holds(\neg\alpha, \mathsf{ff}) & = & \neg\overline{\alpha}^-
\end{array}$$

Definition 12 [Belief Revision in Belnap's four-valued logic] Let ψ and φ be two Belnap formulae and ψ^+ (ψ^-) and φ^+ (φ^-) their translation to classical logic as described above, respectively. Let $\mathcal{A}_B$ be the translation of Belnap's axioms into classical logic and Acc a suitable notion of acceptability of Belnap's theories.

The revision of ψ be φ, in symbols $\psi *_b \varphi$, is defined as

$$\psi *_b \varphi = \{\gamma \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \gamma^+ \text{ and } \neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\gamma^-\}$$

Motivation

The motivation for the above definition is as follows. The translation of any Belnap theory to classical logic is always consistent. What introduces inconsistency into the classical logic translation is the axiom for acceptability. Therefore, without acceptability, the revision of a Belnap theory in classical logic is equivalent to its expansion by the new formula, This is guaranteed by postulates (K*3) and (K*4), and is indeed as it should be, since the object logic is paraconsistent.

To get this correspondence, we need the revision of ψ by φ (without Acc) to be simply $\psi \wedge \varphi$. The correspondence axiom tell us that $\psi \wedge \varphi \rightarrow \gamma$ iff $\mathcal{A}_B \wedge (\psi \wedge \varphi)^+ \vdash \gamma^+$ and $\mathcal{A}_B \wedge \neg(\psi \wedge \varphi)^- \vdash \neg\gamma^-$. Since $\varphi^+ \wedge \mathcal{A}_B$ is consistent with ψ^+, $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B)$ is just $\psi^+ \wedge \varphi^+ \wedge \mathcal{A}_B$ (the same for the negative part). The revision then gives us all $\gamma's$ such that $\psi^+ \wedge \varphi^+ \wedge \mathcal{A}_B \vdash \gamma^+$ and $\neg\psi^- \wedge \neg\varphi^- \wedge \mathcal{A}_B \vdash \neg\gamma^-$. Notice that in the presence of $\mathcal{A}_B$, $\psi^+ \wedge \varphi^+ \leftrightarrow (\psi \wedge \varphi)^+$ and $\neg\psi^- \wedge \neg\varphi^- \leftrightarrow \neg(\psi \wedge \varphi)^-$ (from Ax 3 and Ax 4, respectively, in Definition 9, page 17). Therefore, we get all $\gamma's$ such that $\mathcal{A}_B \wedge (\psi \wedge \varphi)^+ \vdash \gamma^+$ and $\mathcal{A}_B \wedge \neg(\psi \wedge \varphi)^- \vdash \neg\gamma^-$, which is $\{\gamma \mid \psi \wedge \varphi \rightarrow \gamma\}$.

There are in fact only minor differences between the definition above and Definition 2. The extra bits come from the fact that Belnap's translation is more complicated, since we need the negated "holds" to ensure that the other half of the lattice **L4** is also taken into account. The positive holds, checks the values for Both and T, where the negative one checks None and T. F does not need to be checked since it is smaller than all the other values in **L4** anyway.

Discussion of the AGM postulates under Belnap's perspective

(K^*1) $\psi *_b \varphi$ is a belief set

In our terms, this means that the set of Belnap sentences that we receive back from the revision process should be closed under Belnap's consequence relation. In other words, $\gamma \in \psi *_b \varphi$ iff $\psi *_b \varphi \to \gamma$. (only if) This is simple, if $\gamma \in \psi *_b \varphi$, it follows by reflexivity of $\to$ that $\psi *_b \varphi \to \gamma$. (if) Suppose that $\psi *_b \varphi \to \gamma$, we need to show that $\gamma \in \psi *_b \varphi$. But look at the sentences in $\psi *_b \varphi$:

$$\psi *_b \varphi = \{\alpha \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+ \text{ and } \neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\alpha^-\}$$

By the correspondence theorem, if $\psi *_b \varphi \to \gamma$, then

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \vdash \gamma^+$$

and

$$\{\mathcal{A}_B\} \cup \{\neg\alpha^- \mid \neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\alpha^-\} \vdash \neg\gamma^-$$

Notice that

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \subseteq \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$$

since by (K^*2), $\mathcal{A}_B \in \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ and by (K^*1), $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ is closed under classical logical consequence and hence contains all such α^+. By monotonicity of classical logic, it follows that $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \gamma^+$. By a similar argument, we can show that $\neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\gamma^-$ and hence $\gamma \in \psi *_b \varphi$.

(K^*2) $\varphi \in \psi *_b \varphi$

We expect that the set containing the Belnap's sentences obtained from the (classical) revision process includes the revising sentence.

This is simple. Remember that by (K^*2), $\varphi^+ \wedge \mathcal{A}_B \wedge Acc \in \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ and $\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc \in \neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc)$. By ($K^*1$), $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \varphi^+ \wedge \mathcal{A}_B \wedge Acc$ and hence $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \varphi^+$. Similarly, $\neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\varphi^-$. By Definition 12, $\varphi \in \psi *_b \varphi$.

(K^*3) $\psi *_b \varphi \subseteq \mathrm{Cn}_b(\psi \cup \{\varphi\})$

(K^*4) If $\neg\varphi \notin \psi$, then $\mathrm{Cn}_b(\psi \cup \{\varphi\}) \subseteq \psi *_b \varphi$

We have argued in Section 2, page 5, that for non-classical logics, these two postulates have the meaning expressed by ($K^*_{3,4}$) below:

($K^*_{3,4}$) If A is consistent with K, then $K *_a A = \mathrm{Cn}_b(K \cup \{A\})$.

In this case, being "consistent" means *acceptable*. In other words, that acceptability plays no role in the revision process. The equivalence above was shown previously in the motivation for Definition 12.

(K*5) $\psi *_b \varphi = \psi_\perp$ only if φ is unsatisfiable

Similarly, this should have the following interpretation

"$\psi *_b \varphi$ is non-acceptable, only if φ is (itself) non-acceptable."

This can be shown, if we remember that we artificially forced the translation of non-acceptable Belnap theories to be classically inconsistent. This means, for instance, that if φ is not itself non-acceptable, then $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ is classically consistent.

Suppose $\psi *_b \varphi$ is non-acceptable, we must show that φ is non-acceptable. From the assumption, it follows that

$$\psi *_b \varphi \to \gamma \tag{4}$$

and

$$\psi *_b \varphi \to \neg\gamma, \tag{5}$$

for some γ in Belnap's language. By the correspondence theorem, it follows from (4) that

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \vdash \gamma^+$$

Notice that

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \subseteq \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$$

since by (K*2), $\mathcal{A}_B \in \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ and by (K*1), $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$ is closed under classical logical consequence and hence contains all such α^+. Also notice that by the correspondence theorem and (5)

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \vdash \overline{\gamma}^+$$

For the same reason,

$$\{\mathcal{A}_B\} \cup \{\alpha^+ \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+\} \subseteq \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc)$$

and therefore

$$\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \gamma^+ \wedge \overline{\gamma}^+ = holds(\gamma, \mathsf{tt}) \wedge holds(\neg\gamma, tt)$$

Now $\mathcal{A}_B \wedge holds(\neg\gamma, tt) \vdash holds(\gamma, \mathsf{ff})$, the revision operation above is closed and hence

$$\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash holds(\gamma, \mathsf{tt}) \wedge holds(\gamma, \mathsf{ff})$$

By (K*2) and (K*1), $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash Acc$ and is therefore inconsistent, hence φ must be non-acceptable.

(K*6) If $\varphi \equiv_b \gamma$, then $\psi *_b \varphi \equiv_b \psi *_b \gamma$

If two Belnap formulae are equivalent, then so are the results of revising the same belief set by either of the formulae.

In Belnap's logic, $\varphi \equiv_b \gamma$ means $\varphi \to \gamma$ and $\varphi \leftarrow \gamma$. By correspondence we have that $\mathcal{A}_B \vdash \mathit{holds}(\varphi, \mathsf{tt}) \leftrightarrow \mathit{holds}(\gamma, \mathsf{tt})$ and also $\mathcal{A}_B \vdash \neg \mathit{holds}(\varphi, \mathsf{ff}) \leftrightarrow \neg \mathit{holds}(\gamma, \mathsf{ff})$. We show the positive part only. The negative part is similar. It is easy to see that $\vdash (\mathcal{A}_B \wedge \varphi^+) \leftrightarrow (\mathcal{A}_B \wedge \gamma^+)$. By (K*6), $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \equiv \psi^+ *_a (\gamma^+ \wedge \mathcal{A}_B \wedge Acc)$. The same happens with the negative part and therefore, by the definition of revision in Belnap's logic, $\psi *_b \varphi = \psi *_b \gamma$, and hence $\psi *_b \varphi \equiv_b \psi *_b \gamma$.

(K*7) $\quad \psi *_b (\varphi \wedge \gamma) \subseteq \mathrm{Cn}_b((\psi *_b \varphi) \cup \{\gamma\})$

(K*8) $\quad$ If $\neg\gamma \notin \psi *_b \varphi$, then $\mathrm{Cn}_b(\psi *_b \varphi \cup \{\gamma\}) \subseteq \psi *_b (\varphi \wedge \gamma)$

We have argued in Section 2, page 7, that these two postulates have the meaning expressed by ($\mathrm{K}^*_{7,8}$) below:

($\mathrm{K}^*_{7,8}$) $\quad \mathrm{Cn}_b(\psi *_b \varphi \cup \{\gamma\}) = \psi *_b (\varphi \wedge \gamma)$, when γ is consistent with $\psi *_b \varphi$

In this case, being "consistent" means *acceptable*. Acceptability plays therefore no role in the revision process above, and hence $\psi *_b \varphi \cup \{\gamma\}$ is *acceptable*. In first order logic terms, this means that the acceptability axioms is never used in the derivation of translated Belnap formulae.

We sketch the proof as follows. Let us look first at the formulae in the set $\psi *_b \varphi \cup \{\gamma\}$. These are the formulae in the set $\{\alpha \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+$ and $\neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\alpha^-\} \cup \{\gamma\}$. On the other hand, the formulae in the set $\psi *_b (\varphi \wedge \gamma)$ are the same formulae in the set $\{\beta \mid \psi^+ *_a ((\varphi \wedge \gamma)^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \beta^+$ and $\neg\psi^- *_a (\neg(\varphi \wedge \psi)^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\beta^-\}$. Now by the assumption and ($\mathrm{K}^*_{7,8}$) for $*_a$, we have that $\mathrm{Cn}(\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \cup \{\gamma^+\}) = \psi^+ *_a ((\varphi \wedge \gamma)^+ \wedge \mathcal{A}_B \wedge Acc)$. Notice that $\mathcal{A}_B \in \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \cup \{\gamma^+\}$. It is easy to show that $\psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \cup \{\gamma^+\} \vdash \beta^+$ and $\neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \cup \{\neg\gamma^-\} \vdash \neg\beta^-$ iff $\beta \in \mathrm{Cn}_b(\{\alpha \mid \psi^+ *_a (\varphi^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+$ and $\neg\psi^- *_a (\neg\varphi^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\alpha^-\} \cup \{\gamma\})$. Similarly, $\psi^+ *_a ((\varphi \wedge \gamma)^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \beta^+$ and $\neg\psi^- *_a (\neg(\varphi \wedge \gamma)^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\beta^-$ iff $\beta \in \mathrm{Cn}_b(\{\alpha \mid \psi^+ *_a ((\varphi \wedge \gamma)^+ \wedge \mathcal{A}_B \wedge Acc) \vdash \alpha^+$ and $\neg\psi^- *_a (\neg(\varphi \wedge \gamma)^- \wedge \mathcal{A}_B \wedge Acc) \vdash \neg\alpha^-\})$.

6 Conclusions

We have presented a way of exporting an AGM revision process in classical logic to other non-classical logics by translating into classical logic. There are considerable benefits to such a revision by translation over any direct revision in the non-classical logic itself.

1. It is a standard for many non-classical logics to be translated into classical logic. Such translations are done for a variety of reasons:
    - to give the logic a meaning
    - to give semantics to the logic
    - to compare it with other logics
    - to get decidability/undecidability results

- to make use of automated deduction of classical logic

  Adding to this culture a revision capability makes sense.

2. Revision theory is very well developed in classical logic. There are various notions and fine tuning involved and translation into classical logic will not only open a wealth of distinctions for the source logic but also enrich classical logic revision itself with new ideas and problems arising from non-classical logics.

3. From the point of view of classical logic, what we are doing is a *relative revision*. This concept can be defined as follows. Given a theory T of classical logic (for example a theory of linearly ordered Abelian groups[6]), we can talk about a set of sentences Δ being acceptable (i.e., $\mathrm{Cn}(T \cup \Delta)$ being acceptable)[7]. We can talk about revision relative to T of Δ by ψ (denoted by $\Delta *_T \psi$) yielding a new acceptable theory $\Delta' = \Delta *_T \psi$. The case where T arises in connection with a translation from another logic is only one instance of this general relative revision.

We need to distinguish two cases in our studies of relative revision

(a) The notion of acceptability can be easily handled in classical logic.

  This case is simpler to handle and includes the translation from modal logics seen in Section 3.

b) The concept of acceptability is not directly expressible. Here we have problems to overcome, as in the Belnap's logic translations.

We leave the investigations of relative revision for a future paper.

7 Acknowledgments

Dov Gabbay, Odinaldo Rodrigues and Alessandra Russo were supported by the UK EPSRC projects "Temporal Logics of Knowledge and Belief" (GR/K57268), "Data Driven Algorithms" (GR/L91818) and "Managing Inconsistency in Software Engineering" (GR/L55964), respectively.

References

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

[2] C. A. Alchourrón and D. Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.

[6] We are preparing here for the future, when we examine revision in fuzzy logic.

[7] Let $\mathcal{M}$ be a class of models of T. Δ is T-consistent if $\Delta \cup T$ has a model. Δ is $\mathcal{M}$-acceptable if $\Delta \cup T$ has a model in $\mathcal{M}$. Given Δ and ψ, we want the result of the revision of Δ by ψ to be $\mathcal{M}$-acceptable.

[3] C. A. Alchourrón and D. Makinson. The logic of theory change: Safe contraction. *Studia Logica*, 44:405–422, 1985.

[4] N. Belnap. How a computer should think. In G. Ryle, editor, *Contemporary Aspects of Philosophy*, pages 30–56. Oriel Press, 1977.

[5] N. Belnap. A useful four-valued logic. In J. M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-valued Logic*, pages 8–37. D. Reidel, 1977.

[6] M. Dalal. Updates in propositional databases. Technical report, Department of Computer Science, Rutgers University, February 1988. Technical Report DCS-TR-222.

[7] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. Technical Report R-202, Cognitive Science Laboratory, Computer Science Department, University of California, Los Angeles, CA 90024, November 1996.

[8] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.

[9] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. A Bradford Book - The MIT Press, Cambridge, Massachusetts - London, England, 1988.

[10] Peter Gärdenfors, editor. *Belief Revision*, number 29 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1992.

[11] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. *Belief Revision*, pages 183–203, 1992.