Explanations, Belief Revision and Defeasible Reasoning

Marcelo A. Falappa [a] Gabriele Kern-Isberner [b]
Guillermo R. Simari [a]

[a] *Artificial Intelligence Research and Development Laboratory*
Department of Computer Science and Engineering - Universidad Nacional del Sur
Av. Alem 1253 - (B8000CPB) Bahía Blanca - Argentina
E-MAIL: [mfalappa,grs]@cs.uns.edu.ar

[b] *Department of Computer Science - LG Praktische Informatik VIII*
FernUniversitaet Hagen - D-58084 Hagen, Germany
E-MAIL: gabriele.kern-isberner@fernuni-hagen.de

Abstract

We present different constructions for non-prioritized belief revision, that is, belief changes in which the input sentences are not always accepted. First, we present the concept of explanation in a deductive way. Second, we define multiple revision operators with respect to sets of sentences (representing explanations), giving representation theorems. Finally, we relate the formulated operators with argumentative systems and default reasoning frameworks.

1 Introduction

Belief Revision systems are logical frameworks for modelling the dynamics of knowledge. That is, how we modify our beliefs when we receive new information. The main problem arises when that information is inconsistent with the beliefs that represent our epistemic state. For instance, suppose we believe that a Ferrari coupe is the fastest car and then we found out that some Porsche cars are faster than any Ferrari cars. Surely, we need to revise our beliefs in order to accept the new information while preserving as much of the old information as possible.

One of the most controversial properties of the revision operators is success. Success specifies that the new information has primacy over the beliefs of an agent. In this work we propose a kind of non-prioritized revision operator in which the new information is supported by an explanation. Every explanation contains an *explanans* (the beliefs that support a conclusion) and an *explanandum* (the final conclusion). Each explanation is a set of sentences with some restrictions. The operators we propose, which are defined upon belief bases, are an intermediate model between semi-revision operator [Han97] and merge operator [Fuh97]. Moreover, we present a close connection between these new operators and frameworks for default reasoning. The idea is that the beliefs to be deleted in a belief base could be preserved in an alternate set as defeasible rules or assumptions.

There are many different frameworks for belief revision but AGM [AGM85] is the one which has received the most attention. Most others rely on the foundations of AGM. They present an epistemic model (the formalism in which the beliefs will be represented) and then they define different kinds of operators. The basic representation of epistemic states is through belief sets (sets of sentences closed under logical consequence) or belief bases (sets of sentences not necessarily closed). Each operator may be presented in two ways: by giving an explicit construction (algorithm) for the operator, or by giving a set of rationality postulates to be satisfied. Rationality postulates determine constraints that the operators should satisfy. They treat the operators as black boxes; after receiving certain inputs (of new information) we know what the response will be, but not the internal mechanisms used.

The operators for change use selection functions to determine which beliefs will be erased from the epistemic state. Partial meet contractions (AGM framework) are based on a selection among subsets of the original set that do not imply the information to be retracted. The kernel contraction approach is based on a selection among the sentences that imply the information to be retracted. Revision operators can be defined through Levi identity; in order to revise an epistemic state with respect to a sentence α, we contract with respect to $\neg\alpha$ and then expand the new epistemic state with respect to α.

1.1 On the use of explanations

The role of explanations in knowledge representation has been widely studied in [BB93,Gä88,WPFS95,Pag96,PNF94]. We can motivate the use of explanations with an example. Suppose that Michael believes that (α) *all birds fly* and that (β) *Tweety is a bird*. Thus, he will believe that (δ) *Tweety flies*. Then, Johana tells him that *Tweety does not fly*. As a consequence, Michael will have to drop the belief in α or the belief in β forced by having to drop δ. However, it does not seem like a rational attitude to incorporate any external belief without pondering it. Usually, an *intelligent* agent demands an explanation supporting the provided information. Even more so if that information contradicts its own set of beliefs. Being rational, Michael will demand an explanation for the belief $\neg\delta$. For instance, Johana accompanies her contention of *Tweety does not fly* with the sentences *Tweety does not fly because it is a penguin* and *penguins are birds but they do not fly.* Perhaps convinced, Michael would have to check his beliefs in order to determine whether he believes *Tweety flies.*

The main role of an explanation is to rationalize facts. At the base of each explanation rests a why-question [Res70]: "Why does Tweety not fly?", "Why did he say what he did?", "Why is it raining?". We think that a rational agent, before incorporating a new belief that contradicts its knowledge, demands an explanation for the provided information by means of a why-question. Then, if the explanation resists the discussion, the new belief, or its explanation, or both are incorporated into the knowledge.

1.2 The belief revision framework

Since explanations are a major instrument for producing rational belief changes, they should be representable in belief revision theory. Unfortunately, AGM theory [AGM85,Gä88], the dominant framework for belief revision, does not seem to allow for an account of explanations, and the same applies to most other frameworks of belief revision that we are aware of.

The reason for this is that an explanation should be capable of inducing belief in a statement that would not be accepted without the explanation; when faced with a statement α, the epistemic agent does not believe in it, but if an explanation A is provided, then he or she will acquire belief in α. This simple feature cannot be modelled in the AGM framework for the simple reason that it only contains two mechanisms for the receipt of new information -expansion and revision- both of which satisfy the success postulate according to which the input information is always accepted.

In our opinion, a better account of explanation can be obtained with a semi-revision operator (non-prioritized belief revision operator). By this, we mean an operator that sometimes accepts the new information and sometimes rejects it. If the new information is accepted, then deletions from the old information are made if this is necessary to maintain consistency. A wide treatment of non-prioritized revision operators on belief sets can be found in [HFCF01].

We will adopt a propositional language $\mathfrak{L}$ with a complete set of boolean connectives: $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$. Formulæ in $\mathfrak{L}$ will be denoted by lowercase Greek characters: $\alpha, \beta, \delta, \ldots, \omega$. Sets of sentences in $\mathfrak{L}$ will be denoted by uppercase Latin characters: $A, B, C, \ldots, Z$. The symbol $\top$ represents a tautology or *truth.* The symbol $\perp$ represents a contradiction or *falsum.* The characters γ and σ will be reserved to represent selection functions for change operators. We also use a consequence operator Cn. Cn takes sets of sentences in $\mathfrak{L}$ and produces new sets of sentences. The operator Cn satisfies *inclusion* ($A \subseteq Cn(A)$), *iteration* ($Cn(A) = Cn(Cn(A))$), and *monotony* (if $A \subseteq B$ then $Cn(A) \subseteq Cn(B)$). We will assume that the consequence operator includes classical consequences and verifies the standard properties of *supraclassicality* (if α can be derived from A by deduction in classical logic, then $\alpha \in Cn(A)$), *deduction* ($\beta \in Cn(A \cup \{\alpha\})$ if and only if $(\alpha \rightarrow \beta) \in Cn(A)$) and *compactness* (if $\alpha \in Cn(A)$ then $\alpha \in Cn(A')$ for some finite subset A' of A). To simplify notation, we write $Cn(\alpha)$ for $Cn(\{\alpha\})$ where α is any sentence in $\mathfrak{L}$. We also write $\alpha \in Cn(A)$ as $A \vdash \alpha$.

Let K be a set of sentences. As in the AGM framework, we will assume three different epistemic attitudes: *accepted* (whenever $\alpha \in Cn(K)$), *rejected* (whenever $\neg\alpha \in Cn(K)$) and *undetermined* (whenever $\alpha \notin Cn(K)$ and $\neg\alpha \notin Cn(K)$).

2 Explanations in the belief revision framework

In order to present a revision operator based on explanations, we will first define an explanation. An explanation contains two main parts: an *explanans*, that is, the beliefs that support a conclusion, and an *explanandum*, that is, the final conclusion of the explanans. We will use a set of sentences as the explanans and a single sentence as the explanandum.

Definition 1 *The set A is an* explanation *for the sentence α if and only if the following properties are satisfied:*

(1) Deduction*: $A \vdash \alpha$.*
(2) Consistency*: $A \nvdash \perp$.*
(3) Minimality*: If $B \subset A$ then $B \nvdash \alpha$.*
(4) Informational Content*: $Cn(A) \nsubseteq Cn(\alpha)$.*

The relation A explains *α will be noted as $A \rightarrowtail \alpha$.*

Deduction determines that the explanans implies the explanandum. Consistency averts the possibility that a conclusion be derived from an inconsistent set. Minimality establishes that there are no irrelevant beliefs in the explanans. Informational content precludes that the explanandum would imply every sentence in the explanans (for example, $A = \{\alpha \vee \beta, \alpha \vee \neg\beta\}$ is not an explanation for α because $Cn(A) \subseteq Cn(\alpha)$). Moreover, informational content precludes that a single sentence could be an explanation for itself (this means that it is not the case that $\{\alpha\} \rightarrowtail \alpha$ for any sentence α).

In dialogues between two agents it is very common that an agent does not fully accept the information provided by the other. Moreover, it is typical that an agent accepts the new information partially. So, we will define a revision operator of non-prioritized revision in order to capture this behavior. Other work related to partial acceptance was formulated by Fermé and Hansson [FH98].

Now we will present postulates for a revision operator by a *set of sentences*. We will extend the framework in [Han97] to allow for multiple inputs, *i.e.*, for sets of sentences (explanations) as input. Therefore, our operator will be a function that takes us from two sets of sentences to a new set of sentences. We will assume that A is a set of sentences. Let K be a set of sentences and "$\circ$" a revision operator. We propose the following postulates[1]:

Inclusion: $K \circ A \subseteq K \cup A$.
This postulate establishes that, if an agent revises its belief base K with respect to a set A, then its new stock of beliefs will be contained in the union of K and A.

Vacuity: If $K \cup A \nvdash \perp$ then $K \circ A = K \cup A$.
This postulate establishes that if the input set A is consistent with the original beliefs K, then the revised belief base is equal to the union of K and A.

Vacuity 2: If $A \subseteq K$ and $K \nvdash \perp$ then $K \circ A = K$.
This postulate determines that if the input set is already included in a consistent belief base K, then the revised belief base is equal to K.

Weak Success: If $K \cup A \nvdash \perp$ then $A \subseteq K \circ A$.
This postulate says that A is included in the revised belief base whenever A is consistent with K.

Stability: If $A \subseteq K$ and $K \nvdash \perp$ then $A \subseteq K \circ A$.
This postulate says that A is included in the revised belief base whenever A is already included in K and K is consistent.

Consistency: If $A \nvdash \perp$ then $K \circ A \nvdash \perp$.
This postulate is equivalent to consistency postulate of the AGM model, in which the revised set is consistent if the input set is consistent.

Consistency Preservation: If $K \nvdash \perp$ then $K \circ A \nvdash \perp$.
This postulate ensures that the revised belief base is consistent whenever the original belief base is consistent.

Strong Consistency: $K \circ A \nvdash \perp$.
This postulate ensures consistency in the revised belief base.

Core Retainment: If $\alpha \in (K \cup A) \setminus (K \circ A)$ then there is a set H such that $H \subseteq (K \cup A)$, H is consistent but $H \cup \{\alpha\}$ is inconsistent.

[1] *Core Retainment* and *Relevance* have been modified from Hansson's works [Han91,Han96] to be used in our formalism. Similarly, *Congruence* has been modified from Fuhrmann's work [Fuh97].

This postulate expresses the intuition that nothing is removed from the union of the original belief base and the input set unless its removal in some way contributes to making the new belief base consistent.

Relevance: If $\alpha \in (K \cup A) \setminus (K \circ A)$ then there is a set H such that $K \circ A \subseteq H \subseteq (K \cup A)$, H is consistent but $H \cup \{\alpha\}$ is inconsistent.

This postulate is a stronger version of core retainment and we will use it to characterize some kinds of revision operators.

Congruence: If $K \cup A = K \cup B$ then $K \circ A = K \circ B$.

This postulate expresses that if A joined with K is equal to B joined with K then the revision with respect to A is equal to the revision with respect to B.

Fairness: If the condition that $A \nvdash \perp$, $B \nvdash \perp$ and for all $H \subseteq K$ holds that $(H \cup A) \vdash \perp$ if and only if $(H \cup B) \vdash \perp$, then $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$.

This postulate establishes that, if any subset H of K is inconsistent with a consistent set A if and only if it is inconsistent with a consistent set B, then the sentences erased in the respective revisions with respect to A and B are the same.

Reversion: If $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets then $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$.

This postulate establishes that, if $K \cup A$ and $K \cup B$ contain the same minimally inconsistent subsets then the sentences erased in the respective revisions with respect to A and B are the same.

Weak Monotony: If $A \subseteq B$ and $K \cup B \nvdash \perp$ then $K \circ A \subseteq K \circ B$.

This postulate establishes that if a set B contains a subset A and B is consistent with K, then the revision of K with respect to A will be contained in the revision of K with respect to B.

Proposition 2 *Some interesting relations among postulates:*

(1) If "$\circ$" satisfies relevance *then it satisfies* core retainment.
(2) If "$\circ$" satisfies strong consistency *then it satisfies* consistency *and* consistency preservation.
(3) If "$\circ$" satisfies inclusion *and* core retainment *then it satisfies* vacuity.
(4) If "$\circ$" satisfies inclusion *and* reversion *then it satisfies* congruence.
(5) If "$\circ$" satisfies vacuity *then it satisfies* weak success.
(6) If "$\circ$" satisfies inclusion *and* vacuity *then it satisfies* weak monotony.
(7) If "$\circ$" satisfies inclusion *and* fairness *then it satisfies* weak monotony.
(8) If "$\circ$" satisfies vacuity *then it satisfies* vacuity 2.

PROOF.

(1) Straightforward.
(2) Straightforward.

(3) Let $K \cup A \nvdash \perp$. We must show that $K \circ A = K \cup A$. By *inclusion* $K \circ A \subseteq K \cup A$. It remains to show that $K \cup A \subseteq K \circ A$. Suppose, to the contrary, that $K \cup A \nsubseteq K \circ A$. That is, there is some α such that $\alpha \in K \cup A$ but $\alpha \notin K \circ A$. Then $\alpha \in (K \cup A) \setminus (K \circ A)$. By *core retainment* there is a set H such that $H \subseteq K \cup A$, $H \nvdash \perp$ but $H \cup \{\alpha\} \vdash \perp$. Since $\alpha \in K \cup A$ and $H \subseteq K \cup A$ then $H \cup \{\alpha\} \subseteq K \cup A$. Therefore, $K \cup A \vdash \perp$. This contradiction establishes the claim.

(4) Let $K \cup A = K \cup B$. Then $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets. From **reversion** we have that $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$. We need to show that $K \circ A = K \circ B$.

Assume, to the contrary, that $K \circ A \neq K \circ B$. That is, there is a sentence $\alpha \in K \circ A$ and $\alpha \notin K \circ B$. From **inclusion** it follows that $\alpha \in K \cup A$. Since $K \cup A = K \cup B$ then $\alpha \notin (K \cup A) \setminus (K \circ A)$ and $\alpha \in (K \cup B) \setminus (K \circ B)$. This contradiction establishes the claim.

(5) Straightforward.

(6) Let $A \subseteq B$ and $K \cup B \nvdash \perp$. We must show that $K \circ A \subseteq K \circ B$. By *inclusion* we have that $K \circ A \subseteq K \cup A$. Since "$\circ$" satisfies *vacuity* then $K \cup B \nvdash \perp$ implies that $K \circ B = K \cup B$. Since $K \cup A \subseteq K \cup B$ then $K \circ A \subseteq K \circ B$ and we are done.

(7) Let $A \subseteq B$ and $K \cup B \nvdash \perp$. We must show that $K \circ A \subseteq K \circ B$. Assume, to the contrary, that $K \circ A \nsubseteq K \circ B$. Then there is a sentence α such that $\alpha \in K \circ A$ and $\alpha \notin K \circ B$. By *inclusion* if $\alpha \in K \circ A$ then $\alpha \in K \cup A$. Since $A \subseteq B$ then $\alpha \in K \cup B$. Then $\alpha \in K \circ A$, $\alpha \in K \cup A$, $\alpha \notin K \circ B$ and $\alpha \in K \cup B$. That means that $\alpha \notin (K \cup A) \setminus (K \circ A)$ and $\alpha \in (K \cup B) \setminus (K \circ B)$. Therefore $(K \cup A) \setminus (K \circ A) \neq (K \cup B) \setminus (K \circ B)$. From the hypothesis we have that $K \cup A \nvdash \perp$. Therefore, for all $H \subseteq K$ we have that $H \cup A \nvdash \perp$ and $H \cup B \nvdash \perp$. By *fairness* $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$. This contradiction establishes the claim.

(8) Let $A \subseteq K$ and $K \nvdash \perp$. We must show that $K \circ A = K$. Since $A \subseteq K$ and $K \nvdash \perp$ then $K \cup A \nvdash \perp$. By *vacuity* $K \circ A = K \cup A$. Since $A \subseteq K$ then $K \circ A = K \cup A = K$. $\square$

The above properties are proposed as basic requirements for an account of explanation in a belief revision framework. The mechanism of a revision operator by a set of sentences with partial acceptance is:

(1) The input set A is initially accepted.
(2) All possible inconsistencies of $K \cup A$ are removed.

This operator is an operator of *external revision*. The name "external" indicates that the revision process takes place outside of the original set. We can see that there is an intermediate stage in which the epistemic state can be inconsistent.

The operator we will define is an intermediate form between two non prioritized revision operators: semi-revision and merge. Semi-revision is a non prioritized revision operator proposed by Hansson [Han97] that allows the revision of a set K with respect to a single sentence α. On the other hand, a merge operator was presented by Fuhrmann [Fuh97] and it allows the revision of two arbitrary sets of sentences.

2.1 Kernel Revision by a Set of Sentences

The first construction of revision by a set of sentences is based on the concept of a kernel set.

Definition 3 (Hansson [Han93]) *Let K be a set of sentences and α a sentence. Then $K^{\perp\!\!\perp}\alpha$ is the set of all K' such that $K' \in K^{\perp\!\!\perp}\alpha$ if and only if $K' \subseteq K$, $K' \vdash \alpha$, and if $K'' \subset K'$ then $K'' \nvdash \alpha$. The set $K^{\perp\!\!\perp}\alpha$ is called the* kernel set*, and its elements are called the α-*kernels of *K.*

For instance, if $K = \{p, p \rightarrow q, r, r \rightarrow s, r \wedge s \rightarrow q, t \rightarrow u\}$ then the set of q-kernels is equal to $\{\{p, p \rightarrow q\}, \{r, r \rightarrow s, r \wedge s \rightarrow q\}\}$. If $K = \{p, p \rightarrow q\}$ then $K^{\perp\!\!\perp}(p \rightarrow p) = \{\varnothing\}$ because $p \rightarrow p \in Cn(\varnothing)$ and $K^{\perp\!\!\perp}\neg p = \varnothing$ since $K \nvdash \neg p$.

In order to define the operator of revision by a set of sentences we need to use an incision function. This function selects sentences to be removed and it is called incision function because it makes an incision in every $\perp$-kernel. However, this function is not only applied to K. It is also applied to supersets of K. Therefore, we need an external incision function for K.

Definition 4 *Let K be a set of sentences. An* external incision function *for K is a function "σ"($\sigma : \mathbf{2^{2^{\mathfrak{L}}}} \Rightarrow \mathbf{2^{\mathfrak{L}}}$) such that for any set $A \subseteq \mathfrak{L}$, the following hold:*

1) $\sigma((K \cup A)^{\perp\!\!\perp}\perp) \subseteq \cup((K \cup A)^{\perp\!\!\perp}\perp)$.

2) If $X \in (K \cup A)^{\perp\!\!\perp}\perp$ and $X \neq \varnothing$ then $(X \cap \sigma((K \cup A^{\perp\!\!\perp}\perp)) \neq \varnothing$.

The limit case in which $(K \cup A)^{\perp\!\!\perp}\perp = \varnothing$ then $\sigma((K \cup A)^{\perp\!\!\perp}\perp) = \varnothing$.

For instance, taking $K = \{t, u, r, r \rightarrow s\}$ and $A = \{\neg t, p, p \rightarrow \neg s\}$ then $K \cup A = \{t, u, r, r \rightarrow s, \neg t, p, p \rightarrow \neg s\}$, $(K \cup A)^{\perp\!\!\perp}\perp = \{\{r, r \rightarrow s, p, p \rightarrow \neg s\}, \{t, \neg t\}\}$, and some possible results of $\sigma((K \cup A)^{\perp\!\!\perp}\perp)$ are $\{p, t\}$, $\{p, \neg t\}$ and $\{p \rightarrow \neg s, t\}$.

Formally, we define the kernel revision by a set of sentences as follows.

Definition 5 *Let K and A be sets of sentences and "σ" an external incision function for K. The operator "$\circ$" of* kernel revision a by set of sentences

$(\circ : \mathbf{2}^{\mathcal{L}} \times \mathbf{2}^{\mathcal{L}} \Rightarrow \mathbf{2}^{\mathcal{L}})$ *is defined as* $K \circ A = (K \cup A) \setminus \sigma((K \cup A)^{\perp\!\!\perp}\bot)$.

The mechanism of this operator is to add A to K and then eliminate from the result all possible inconsistency by means of an incision function that makes a "cut" over each minimally inconsistent subset of $K \cup A$. Since this operator uses an incision function and the set of $\bot$-kernels, we call it kernel revision by a set of sentences.

An axiomatic characterization can now be given for this kind of operator.

Theorem 6 *Let K be a belief base. The operator "$\circ$" is a* kernel revision by a set of sentences *if and only if it satisfies* inclusion*,* strong consistency*,* core retainment *and* reversion*.*

PROOF.

[Construction to Postulates] Let "$\circ_\sigma$" be a kernel revision by a set of sentences for K. We must show that "$\circ_\sigma$" satisfies the postulates enumerated in the theorem. Let $K \circ_\sigma A = (K \cup A) \setminus (\sigma((K \cup A)^{\perp\!\!\perp}\bot))$.

Inclusion: Straightforward from the definition.

Strong Consistency: Since all sets in $(K \cup A)^{\perp\!\!\perp}\bot$ are minimally inconsistent, and σ cuts every set in it, then $(K \cup A) \setminus \sigma((K \cup A)^{\perp\!\!\perp}\bot)$ is consistent.

Core Retainment: Suppose that $\alpha \in (K \cup A) \setminus (K \circ_\sigma A)$. That is, $\alpha \in K \cup A$ and $\alpha \notin K \circ_\sigma A$. Then $\alpha \in \sigma((K \cup A)^{\perp\!\!\perp}\bot)$. Since $\sigma((K \cup A)^{\perp\!\!\perp}\bot) \subseteq \cup((K \cup A)^{\perp\!\!\perp}\bot)$ there is some X such that $\alpha \in X$ and $X \in (K \cup A)^{\perp\!\!\perp}\bot$. Let $Y = X \setminus \{\alpha\}$. Then there is some Y such that $Y \subseteq (K \cup A)$, $Y \nvdash \bot$ but $Y \cup \{\alpha\} \vdash \bot$. Therefore, core retainment is satisfied.

Reversion: Suppose that $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets. That means that $(K \cup A)^{\perp\!\!\perp}\bot = (K \cup B)^{\perp\!\!\perp}\bot$. Since σ is a well defined function then $\sigma((K \cup A)^{\perp\!\!\perp}\bot) = \sigma((K \cup A)^{\perp\!\!\perp}\bot)$. We need to show that $(K \cup A) \setminus (K \circ_\sigma A) = (K \cup B) \setminus (K \circ_\sigma B)$.

$\subseteq$) If $\alpha \in (K \cup A) \setminus (K \circ_\sigma A)$ then, by definition of "$\circ$", $\alpha \in \sigma((K \cup A)^{\perp\!\!\perp}\bot)$. Since $\sigma((K \cup B)^{\perp\!\!\perp}\bot) = \sigma((K \cup A)^{\perp\!\!\perp}\bot)$ then $\alpha \in K \cup B$ and $\alpha \notin K \circ_\sigma B$. Therefore, $(K \cup A) \setminus (K \circ_\sigma A) \subseteq (K \cup B) \setminus (K \circ_\sigma B)$.

$\supseteq$) If $\alpha \in (K \cup B) \setminus (K \circ_\sigma B)$ then, by definition of "$\circ$", $\alpha \in \sigma((K \cup B)^{\perp\!\!\perp}\bot)$. Since $\sigma((K \cup B)^{\perp\!\!\perp}\bot) = \sigma((K \cup A)^{\perp\!\!\perp}\bot)$ then $\alpha \in K \cup A$ and $\alpha \notin K \circ_\sigma A$. Therefore, $(K \cup B) \setminus (K \circ_\sigma B) \subseteq (K \cup A) \setminus (K \circ_\sigma A)$.

[Postulates to Construction] We need to show that if an operator satisfies the enumerated postulates then it is possible to build an operator in the way specified in the theorem. Let "σ" be a function such that, for every pair of sets K and A, it holds that:

$$\sigma((K \cup A)^{\perp\!\!\perp}\bot) = \{\alpha : \alpha \in (K \cup A) \setminus (K \circ A)\}$$

We must show:

Part A.

(1) "σ" is a well defined function.

That is, if A and B are sets of sentences such that $(K \cup A)^{\perp\!\!\perp}\bot = (K \cup B)^{\perp\!\!\perp}\bot$, we must show that $\sigma((K \cup A)^{\perp\!\!\perp}\bot) = \sigma((K \cup B)^{\perp\!\!\perp}\bot)$. From the hypothesis we have that $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets. It follows from **reversion** that $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$. Therefore:

$$\begin{aligned}\sigma((K \cup A)^{\perp\!\!\perp}\bot) &= \{\alpha : \alpha \in (K \cup A) \setminus (K \circ A)\} \\ &= \{\alpha : \alpha \in (K \cup B) \setminus (K \circ B)\} \\ &= \sigma((K \cup B)^{\perp\!\!\perp}\bot)\end{aligned}$$

Therefore, "σ" is well defined.

(2) $\sigma((K \cup A)^{\perp\!\!\perp}\bot) \subseteq \cup((K \cup A)^{\perp\!\!\perp}\bot)$.

Let $\alpha \in \sigma((K \cup A)^{\perp\!\!\perp}\bot)$. Then $\alpha \in (K \cup A) \setminus (K \circ A)$. Due to **core retainment** there is some H such that $H \subseteq (K \cup A)$, $H \nvdash \bot$ but $H \cup \{\alpha\} \vdash \bot$. Since $\alpha \in K \cup A$ then there is a $\bot$-kernel K' in $(K \cup A)$ (*i.e.*, there is a minimally inconsistent subset of $K \cup A$) such that $K' \subseteq H \cup \{\alpha\}$ and $\alpha \in K'$. Therefore, $\alpha \in \cup((K \cup A)^{\perp\!\!\perp}\bot)$.

(3) If $X \in (K \cup A)^{\perp\!\!\perp}\bot$ then $(X \cap \sigma((K \cup A)^{\perp\!\!\perp}\bot)) \neq \varnothing$.

Let $X \in ((K \cup A)^{\perp\!\!\perp}\bot)$. We need to show that $X \cap \sigma((K \cup A)^{\perp\!\!\perp}\bot) \neq \varnothing$. Due to **strong consistency** $K \circ A \nvdash \bot$. Since $X \vdash \bot$ we may conclude that $X \nsubseteq K \circ A$. This means that there is some β such that $\beta \in X$ and $\beta \notin K \circ A$. Since $X \subseteq (K \cup A)$ then $\beta \in (K \cup A) \setminus (K \circ A)$, *i.e.*, $\beta \in \sigma((K \cup A)^{\perp\!\!\perp}\bot)$. So $\beta \in (X \cap \sigma((K \cup A)^{\perp\!\!\perp}\bot))$. Therefore, $(X \cap \sigma((K \cup A)^{\perp\!\!\perp}\bot)) \neq \varnothing$.

Part B: "$\circ_\sigma$" is equal to "$\circ$".

Due to **inclusion** and from the definition of $\sigma((K \cup A)^{\perp\!\!\perp}\bot)$ we conclude that $K \circ A = K \circ_\sigma A$. $\square$

2.2 Partial Meet Revision by a Set of Sentences

The second construction of revision by a set of sentences is based on the concept of a remainder set.

Definition 7 (Alchourrón, Gärdenfors & Makinson [AGM85]) *Let K be a set of sentences and α a sentence. Then $K^\perp\alpha$ is the set of all K' such that $K' \in K^\perp\alpha$ if and only if $K' \subseteq K$, $K' \nvdash \alpha$ and if $K' \subset K'' \subseteq K$ then $K'' \vdash \alpha$. The set $K^\perp\alpha$ is called the* remainder set *of K with respect to α, and its elements are called the α*-remainders of *K.*

For instance, if $K = \{p, p \to q, r, r \to s, r \wedge s \to q, t \to u\}$ then the set of q-remainders is $\{\{p \to q, r \to s, r \wedge s \to q, t \to u\}, \{p, r \to s, r \wedge s \to q, t \to u\}, \{p \to q, r, r \wedge s \to q, t \to u\}, \{p, r, r \wedge s \to q, t \to u\}\}$. The set of v-remainders of K is equal to $\{K\}$ since $K \nvdash v$. The set of $(p \rightarrow p)$-remainders of K is $\varnothing$ because $p \rightarrow p \in Cn(\varnothing)$ and there is no subset of K failing to imply $p \rightarrow p$.

In order to define the partial meet version of this operator, we need an external selection function, that is, a selection function to be applied over supersets of K.

Definition 8 *Let K be a set of sentences. An* external selection function for *K is a function "γ" ($\gamma : \mathbf{2^{2^{\mathfrak{L}}}} \Rightarrow \mathbf{2^{2^{\mathfrak{L}}}}$) such that for any set $A \subseteq \mathfrak{L}$, it holds that:*

1) $\gamma((K \cup A)^{\perp}\perp) \subseteq (K \cup A)^{\perp}\perp$.

2) $\gamma((K \cup A)^{\perp}\perp) \neq \varnothing$.

Since every set H contains a consistent subset then $H^{\perp}\perp$ is always non-empty. For instance, if $K = \{p, q, r\}$ and $A = \{\neg p, \neg q\}$ then $K \cup A = \{p, q, r, \neg p, \neg q\}$, $(K \cup A)^{\perp}\perp = \{\{p, q, r\}, \{\neg p, q, r\}, \{p, \neg q, r\}, \{\neg p, \neg q, r\}\}$ and some possible results of $\gamma((K \cup A)^{\perp}\perp)$ are $\{\{\neg p, \neg q, r\}\}$, $\{\{\neg p, q, r\}\}$, $\{\{p, q, r\}, \{\neg p, q, r\}\}$ and $\{\{p, q, r\}, \{\neg p, q, r\}, \{p, \neg q, r\}\}$.

Definition 9 *Let K be a set of sentences and γ an external selection function for K. Then γ is an* equitable selection function for K *if $(K \cup A)^{\perp\!\!\perp}\perp = (K \cup B)^{\perp\!\!\perp}\perp$ implies that $(K \cup A) \setminus \cap\gamma((K \cup A)^{\perp}\perp) = (K \cup B) \setminus \cap\gamma((K \cup B)^{\perp}\perp)$.*

The intuition behind this definition is that, if the set of minimally inconsistent subsets of $K \cup A$ is equal to the set of minimally inconsistent subsets of $K \cup B$ then α is erased in the selection of $\perp$-remainders of $K \cup A$ if and only if it is erased in the selection of $\perp$-remainders of $K \cup B$.

For example, let $K = \{a, b, \neg c\}$, $A = \{\neg b, c, d, e\}$ and $B = \{\neg b, c, f\}$. Then:

$$(K \cup A)^{\perp}\perp = \{\{a, b, c, d, e\}, \{a, \neg b, c, d, e\}, \{a, b, \neg c, d, e\}, \{a, \neg b, \neg c, d, e\}\}$$

$$(K \cup B)^{\perp}\perp = \{\{a, b, c, f\}, \{a, \neg b, c, f\}, \{a, b, \neg c, f\}, \{a, \neg b, \neg c, f\}\}$$

We have that $(K \cup A)^{\perp\!\!\perp}\perp = (K \cup A)^{\perp\!\!\perp}\perp = \{\{b, \neg b\}, \{c, \neg c\}\}$. Suppose that $\gamma((K \cup A)^{\perp}\perp) = \{\{a, b, c, d, e\}, \{a, \neg b, c, d, e\}\}$. That is, γ selects only $\perp$-remainders containing c. If γ is an equitable selection function $\gamma((K \cup B)^{\perp}\perp)$ must be equal to $\{\{a, b, c, f\}, \{a, \neg b, c, f\}\}$.

Formally, we define the operator of partial meet revision by a set of sentences as follows.

Definition 10 *Let K and A be sets of sentences and "γ" an equitable selection function for K. The operator "$\circ$" of* partial meet revision by a set of sentences *($\circ : \mathbf{2^{\mathfrak{L}}} \times \mathbf{2^{\mathfrak{L}}} \Rightarrow \mathbf{2^{\mathfrak{L}}}$) is defined as $K \circ A = \cap\gamma((K \cup A)^{\perp}\perp)$.*

The mechanism of this operator is to add A to K and then eliminate from the result all possible inconsistencies by means of an equitable selection function that makes a choice among the maximally consistent subsets of $K \cup A$ and intersect them. Since this operator uses a selection function and the remainder set, we call it partial meet revision by a set of sentences.

The following lemma will be used in the representation theorem of partial meet revision by a set of sentences.

Lemma 11 *If $A^{\perp}\perp = B^{\perp}\perp$ then $A = B$.*

PROOF.

$\subseteq$) Let $A^{\perp}\perp = B^{\perp}\perp$ and let $\alpha \in A$. Then there is an $X \in A^{\perp}\perp$ with $\alpha \in X$. From the presupposition, $X \in B^{\perp}\perp$ too, in particular, $X \subseteq B$; so also $\alpha \in B$. Therefore $A \subseteq B$.

$\supseteq$) Let $B^{\perp}\perp = A^{\perp}\perp$ and let $\alpha \in B$. Then there is an $X \in B^{\perp}\perp$ with $\alpha \in X$. From the presupposition, $X \in A^{\perp}\perp$ too, in particular, $X \subseteq A$; so also $\alpha \in A$. Therefore $B \subseteq A$. $\square$

Thus, an axiomatic characterization can now be given for an operator of partial meet revision by a set of sentences.

Theorem 12 *Let K be a belief base. The operator "$\circ$" is a* partial meet revision by a set of sentences *if and only if it satisfies* inclusion*,* strong consistency*,* relevance *and* reversion*.*

PROOF.

[CONSTRUCTION TO POSTULATES] Let "$\circ_\gamma$" be a partial meet revision by a set of sentence for K. We must show that "$\circ_\gamma$" satisfies the postulates enumerated in the theorem. Let $K \circ_\gamma A = \cap\gamma((K \cup A)^{\perp}\perp)$.

Inclusion: Straightforward from the definition.

Strong Consistency: Since all sets in $(K \cup A)^{\perp}\perp$ are consistent, so is their intersection.

Relevance: Suppose that $\alpha \in (K \cup A) \setminus (K \circ_\gamma A)$. That is, $\alpha \in K \cup A$ and $\alpha \notin K \circ_\gamma A$. Then $\alpha \notin \cap\, \gamma((K \cup A)^{\perp}\perp)$. Since $\gamma((K \cup A)^{\perp}\perp) \subseteq (K \cup A)^{\perp}\perp$ there is some X such that $\alpha \notin X$ and $X \in \gamma((K \cup A)^{\perp}\perp)$. Since $\cap\gamma((K \cup A)^{\perp}\perp) \subseteq X$ then $K \circ_\gamma A \subseteq X$. Since $\alpha \notin X$ then $K \circ_\gamma A \subseteq X \subseteq K \cup A$, $X \nvdash \perp$ but $X \cup \{\alpha\} \vdash \perp$. Therefore, relevance is satisfied.

Reversion: Suppose that $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets, that is, $(K \cup A)^{\perp\!\!\perp}\perp = (K \cup B)^{\perp\!\!\perp}\perp$. We need to show that $(K \cup A) \setminus (K \circ_\gamma A) = (K \cup B) \setminus (K \circ_\gamma B)$. Straightforward since γ is an equitable selection function.

[Postulates to Construction] We will show that if an operator satisfies the enumerated postulates then it is possible to build an operator in the way specified in the theorem. Let "γ" be a function such that, for all pair of sets K and A, it holds that:

$$\gamma((K \cup A)^{\perp}\perp) = \{X \in (K \cup A)^{\perp}\perp : K \circ A \subseteq X\}$$

We must show that:

Part A: "$\circ_\gamma$" is equal to "$\circ$", *i.e.*, $\cap\gamma((K \cup A)^{\perp}\perp) = K \circ A$.

($\supseteq$) It follows from the definition.

($\subseteq$) Let $\alpha \notin K \circ A$. We must prove that $\alpha \notin \cap\gamma((K \cup A)^{\perp}\perp)$. That is, we need to find some $X \in (K \cup A)^{\perp}\perp$ such that $\alpha \notin X$. We have two cases:

(1) $\alpha \in K \cup A$: by **relevance** we have that there is some H such that $K \circ A \subseteq H \subseteq K \cup A$, $H \nvdash \perp$ and $H \cup \{\alpha\} \vdash \perp$. From this we have that $H \nvdash \alpha$. It is clear that we may extend the set H to a maximally consistent set H' such that $H' \in (K \cup A)^{\perp}\perp$ and $\alpha \notin H'$. Since $K \circ A \subseteq H'$ then $H' \in \gamma((K \cup A)^{\perp}\perp)$.

(2) $\alpha \notin K \cup A$: then no set in $(K \cup A)^{\perp}\perp$ will contain α.

Part B.

(1) "γ" is a well defined function.

Let A and B be sets of sentences such that $(K \cup A)^{\perp}\perp = (K \cup B)^{\perp}\perp$. We must show that $\gamma((K \cup A)^{\perp}\perp) = \gamma((K \cup B)^{\perp}\perp)$. If $(K \cup A)^{\perp}\perp = (K \cup B)^{\perp}\perp$ then it follows from Lemma 11 that $K \cup A = K \cup B$. From Proposition 2 it follows that if **inclusion** and **reversion** hold then congruence holds. Therefore, $K \circ A = K \circ B$ and:

$$\begin{aligned}\gamma((K \cup A)^{\perp}\perp) &= \{X \in (K \cup A)^{\perp}\perp : K \circ A \subseteq X\}\\ &= \{X \in (K \cup B)^{\perp}\perp : K \circ B \subseteq X\}\\ &= \gamma((K \cup B)^{\perp}\perp)\end{aligned}$$

That means that the function "γ" is well defined.

(2) "γ" is an equitable selection function.

First we will show that γ is an external selection function. That is to say that $\varnothing \neq \gamma((K \cup A)^{\perp}\perp) \subseteq (K \cup A)^{\perp}\perp$. By **inclusion** $(K \circ A) \subseteq (K \cup A)$. Due to **strong consistency** $K \circ A \nvdash \perp$ and there is a subset of $K \cup A$ which is consistent. Hence, there must exist a set H between $K \circ A$ and $K \cup A$ which is maximally consistent. Then $H \in (K \cup A)^{\perp}\perp$ and $K \circ A \subseteq H$. Therefore, there exists an $H \in \gamma((K \cup A)^{\perp}\perp)$ and γ is an external selection function.

It remains to show that γ is an equitable selection function. Suppose that $K \cup A$ and $K \cup B$ have the same minimally inconsistent subsets. It follows from **reversion** that $(K \cup A) \setminus (K \circ A) = (K \cup B) \setminus (K \circ B)$. It follows from Part A that $K \circ A = \cap\gamma((K \cup A)^{\perp}\perp)$ and $K \circ B = \cap\gamma((K \cup B)^{\perp}\perp)$. Therefore γ is an equitable selection function. $\square$

Since *relevance* implies *core retainment* the following corollary is trivially shown.

Corollary 13 *Each* partial meet revision by a set of sentences *operator is a* kernel revision by a set of sentences *operator.*

3 Relating Revisions and Explanations

In Section 2 we have introduced postulates for explanans (represented by a set of sentences). Now we will present postulates that relate explanans to the corresponding explanandum. Let K be a belief base, "$\circ$" a revision operator by a set of sentences for K, A and B explanans, and α a sentence of the language.

Explanans Inclusion: If $A \rightarrowtail \alpha$ and $A \subseteq K \circ A$ then $K \circ A \vdash \alpha$.
This postulate establishes that if an agent receives an explanation for a sentence and his belief base contains this explanation then the revised belief base derives the explanandum.

Weak Success 2: If $A \rightarrowtail \alpha$ and $K \cup A \nvdash \bot$ then $K \circ A \vdash \alpha$.
This postulate expresses that, if an agent receives an explanation for some sentence α and the sentences of the explanans are not rejected, then the explanandum will be derived in the revised belief base.

Constrained Success: If $A \rightarrowtail \alpha$ and $K \nvdash \neg\alpha$ then $K \circ A \vdash \alpha$.
This postulate establishes that if an agent receives an explanation for some sentence α not rejected in the original belief base then the explanandum will be accepted in the revised belief base.

Expansion: If $A \rightarrowtail \alpha$ and $K \vdash \alpha$ then $K \circ A \vdash \alpha$.
This postulate says that if an agent accepts a sentence α and then receives a new explanation for it he/she will continue accepting the explained sentence.

It is interesting to note that, since explanations satisfy deduction, the acceptance of the explanans forces the acceptance of the explanandum in the revised set. The following proposition establishes this important relation.

Proposition 14 *If "$\circ$" is a revision operator by a set of sentences then it satisfies* explanans inclusion*.*

PROOF. Suppose that we are revising (in kernel or partial meet mode) K by A and $A \rightarrowtail \alpha$. Since "$\rightarrowtail$" satisfies deduction then $A \vdash \alpha$. If $A \subseteq K \circ A$ then $K \circ A \vdash \alpha$.

There is an important fact to remark regarding the degree of acceptance of the explanans and the explanandum. While the explanans can be explicitly included in the revised set, the explanandum may be inferred from it without actually being included. This difference in the degree of acceptance is motivated in the epistemic model adopted here which is based on sets of sentences not necessarily closed as in the AGM model. For this reason, in each belief base we will have two types of beliefs: basic or explicit belief, and inferred or implicit beliefs.

The following proposition shows that the properties of *constrained success* and *expansion* can not be expected to hold in general.

Proposition 15 *If "$\circ$" is an operator of revision by a set of sentences then in general it does not satisfy neither* constrained success *nor* expansion.

PROOF. Let us consider an operation of partial meet revision by a set of sentences since it is always a kernel revision by a set of sentences (Corollary 13).

Constrained Success. Let p, q, r, s and t be logically independent propositions. Let $K = \{p, s, p \wedge s \rightarrow \neg q, s \rightarrow u\}$ and $A = \{p, p \rightarrow q, q \rightarrow r\}$ an explanation for r. It is clear that $K \nvdash \neg r$. We will make a partial meet revision by a set of sentences of K with respect to A. That is, we need to make a selection among the best maximally consistent subsets of $K \cup A$: $(K \cup A)^{\perp}\bot = \{K_1, K_2, K_3\}$ where:

$K_1 = \{p, s, p \wedge s \rightarrow \neg q, s \rightarrow u, q \rightarrow r\}$.
$K_2 = \{p, s, s \rightarrow u, p \rightarrow q, q \rightarrow r\}$.
$K_3 = \{s, p \wedge s \rightarrow \neg q, s \rightarrow u, p \rightarrow q, q \rightarrow r\}$.

Suppose that K_1 and K_2 are the preferred sets. Then the outcome of the partial meet revision by a set of sentences is $K \circ A = \{p, s, s \rightarrow u, q \rightarrow r\}$. It is clear that $K \circ A \nvdash r$.

Expansion. Let p, q and r be logically independent propositions. Let $K = \{p, p \rightarrow q, r\}$ and $A = \{\neg r, \neg r \rightarrow q\}$. It is clear that $K \vdash q$ and $A \rightarrowtail q$. We need to make a selection among the maximally consistent subsets of $K \cup A$. $(K \cup A)^{\perp}\bot = \{K_1, K_2\}$ where $K_1 = \{p, p \rightarrow q, r, \neg r \rightarrow q\}$ and $K_2 = \{p, p \rightarrow q, \neg r, \neg r \rightarrow q\}$. If we select both sets then $K \circ A$ is $\{p, r, \neg r \rightarrow q\}$, which does not imply q. $\square$

4 Extending the representation language

Now we will present some applications of the operator of revision by a set of sentences. We will assume a representation language that is more expressive than a propositional one. Most frameworks used and defined for belief revision use mainly a propositional language or an extension of it. We will define a language $\mathfrak{L}^+$ which is a subset of a first order logic.

Let $\mathfrak{L}^+$ be the extended knowledge representation language with the same logic connectives used in $\mathfrak{L}$. This language is defined recursively by means of the following BNF grammar:

term ::= variable | constant
list-of-terms ::= term | term "," list-of-terms
wff-atomic ::= predicate "(" list-of-terms ")"
wff-free ::= wff-atomic | "¬" wff-free | wff-free "→" wff-free
wff-free "∧" wff-free | wff-free "∨" wff-free
wff ::= "(" "∀" variable ")" wff | wff-free

Our extended language is first order without functional symbols and without explicit existential quantifiers. The symbols between quotation marks are assumed as symbols in the object language (*i.e.*, they are not meta-symbols). All sentences will be *closed*, that is, each occurrence of any variable is bound to a (universal) quantifier. On the other hand, an occurrence of a variable is *free* if it is not within the scope of any quantifier. An occurrence of a variable is *bounded* if it is within the scope of a quantifier. A *ground* sentence is a sentence without variables. An atomic wff is a *positive literal* and a negated atomic wff is a *negative literal*.

To make a distinction among predicate symbols, constants and variables we will use the Prolog notation [SS96], where predicates and constants are character strings beginning with lowercase letters whereas variables are character strings beginning with uppercase letters.

4.1 Different kinds of beliefs

If we use a propositional language, all beliefs have the same status.[2] Every belief is a symbolic notation that represents knowledge about the real world. However, with a propositional language we cannot make a distinction among

[2] The status is different and independent of other measures such as epistemic entrenchment, plausibility, surprise value, acceptance degree or probability.

objects, functions and relations between objects. Moreover, we cannot determine if a sentence is referred to an object or a collection of objects.

Our proposal is to represent the knowledge with a richer language than the propositional one, with the final goal of distinguishing between two kinds of beliefs:

Particular Beliefs: These beliefs will be mainly represented by ground facts such as $bird(tweety)$, $car(porsche)$, $promoter(bill)$ or $greater(3,2)$.

General Beliefs: These beliefs are referred to collections of objects and they will be general rules such as closed material implications. For instance, $(\forall X)(bird(X) \rightarrow flies(X))$, $(\forall X)(quaker(X) \rightarrow pacifist(X))$.

This distinction is context dependent and it could be modified in different frameworks. Each belief base K has the form $K_{\mathrm{P}} \cup K_{\mathrm{G}}$ where $K_{\mathrm{P}} \cap K_{\mathrm{G}} = \varnothing$. K_{P} is the set of particular sentences whereas K_{G} is the set of general sentences. The same assumption will be made in the explanations since they will contain particular and general sentences of the language.

Example 16 *Let $K = K_{\mathrm{P}} \cup K_{\mathrm{G}}$ be a belief base where:*

$K_{\mathrm{P}} = \{ostr(jim), ostr(tom)\}$

$K_{\mathrm{G}} = \{(\forall X)(ostr(X) \rightarrow bird(X)), (\forall X)(bird(X) \rightarrow flies(X))\}$

The ground logical consequences of this set are:

$$\{ostr(jim), ostr(tom), bird(jim), bird(tom), flies(jim), flies(tom)\}$$

Note that the individuals referenced in this set are relevant, that is, all individuals are mentioned in the belief base. Suppose we receive the following explanation for $\neg flies(jim)$: $A = \{ostr(jim), (\forall X)(ostr(X) \rightarrow \neg flies(X))\}$.

We need to find the minimally inconsistent sets of $K \cup A$:

(1) $\{ostr(jim), (\forall X)(ostr(X) \rightarrow bird(X)),$
$(\forall X)(bird(X) \rightarrow flies(X)), (\forall X)(ostr(X) \rightarrow \neg flies(X))\}$

(2) $\{ostr(tom), (\forall X)(ostr(X) \rightarrow bird(X)),$
$(\forall X)(bird(X) \rightarrow flies(X)), (\forall X)(ostr(X) \rightarrow \neg flies(X))\}$

Suppose we are making a kernel revision by a set of sentences. For that, we need an incision function to make a cut upon every set. That is, we must decide which beliefs must be given up in the revision process. A possible policy could be to discard particular beliefs; on the other hand, we could discard general beliefs. If we choose the latter option we must decide to give up at least one belief in the set:

$\{(\forall X) ostr(X) \rightarrow bird(X), (\forall X) bird(X) \rightarrow \mathit{flies}(X), (\forall X) ostr(X) \rightarrow \neg \mathit{flies}(X)\}$

If we use the notion of specificity[3] *[Poo85,SL92] we could keep the sentence:*

$$(\forall X)(ostr(X) \rightarrow \neg \mathit{flies}(X))$$

and eliminate at least one sentence of the remaining set of sentences. Suppose we give up the sentence $(\forall X)(bird(X) \rightarrow \mathit{flies}(X))$ *That is,* $\sigma((K \cup A)^{\perp\!\!\perp}\bot) = \{(\forall X)(bird(X) \rightarrow \mathit{flies}(X))\}$. *The revised belief base, noted* $K \circ_\sigma A$, *will be composed of the following subsets:*

$K'_{\mathrm{P}} = \{ostr(jim), ostr(tom)\}$

$K'_{\mathrm{G}} = \{(\forall X)(ostr(X) \rightarrow bird(X)), (\forall X)(ostr(X) \rightarrow \neg \mathit{flies}(X))\}$

The ground logical consequences of the revised belief base are:

$$\{ostr(jim), ostr(tom), bird(jim), bird(tom), \neg \mathit{flies}(jim), \neg \mathit{flies}(tom)\}$$

Example 17 *Let* $K = K_{\mathrm{P}} \cup K_{\mathrm{G}}$ *be a belief base such that:*

$K_{\mathrm{P}} = \{bird(tweety), peng(opus)\}$

$K_{\mathrm{G}} = \{(\forall X)(peng(X) \rightarrow bird(X)), (\forall X)(bird(X) \rightarrow \mathit{flies}(X))\}$

The ground logical consequences of K *are:*

$$\{bird(tweety), peng(opus), \mathit{flies}(tweety), bird(opus), \mathit{flies}(opus)\}$$

Suppose we receive the following explanation for $\neg \mathit{flies}(opus)$*:*

$$A = \{bird(opus), peng(opus), (\forall X)(bird(X) \wedge peng(X) \rightarrow \neg \mathit{flies}(X))\}$$

Suppose we are making a kernel revision by a set of sentences. We need to find the minimally inconsistent sets of $K \cup A$. *The sets in these conditions are:*

(1) $\{bird(opus), peng(opus), (\forall X)(bird(X) \wedge peng(X) \rightarrow \neg \mathit{flies}(X))\} \cup \{(\forall X)(bird(X) \rightarrow \mathit{flies}(X))\}$.
(2) $\{peng(opus), (\forall X)(peng(X) \rightarrow bird(X)), (\forall X)(bird(X) \rightarrow \mathit{flies}(X))\} \cup \{(\forall X)(bird(X) \wedge peng(X) \rightarrow \neg \mathit{flies}(X))\}$.

[3] Informally, $\{a \wedge b \rightarrow c\}$ (based on the facts a and b) is more specific than $\{a \rightarrow \neg c\}$ (just based on the fact a). On the other hand, $\{p \rightarrow r\}$ (based on p and one rule) is more specific than $\{p \rightarrow q, q \rightarrow \neg r\}$ (based on p and two rules). A formal definition of specifity can be found in Section 6.

If we choose to discard general beliefs, we must give up at least one sentence of every set:

- $\{(\forall X)(bird(X) \wedge peng(X) \rightarrow \neg flies(X)), (\forall X)(bird(X) \rightarrow flies(X))\}$.
- $\{(\forall X)(peng(X) \rightarrow bird(X)), (\forall X)(bird(X) \wedge peng(X) \rightarrow \neg flies(X))\} \cup \{(\forall X)(bird(X) \rightarrow flies(X))\}$.

Since $(\forall X)(bird(X) \rightarrow flies(X))$ *is a common member of these two sets, we could discard it. That is,* $\sigma((K \cup A)^{\perp\!\!\perp}\bot) = \{(\forall X)(bird(X) \rightarrow flies(X))\}$*. The revised belief base, noted* $K \circ_\sigma A$*, will be composed of the following sets:*

$K'_{\mathrm{P}} = \{bird(tweety), bird(opus), peng(opus)\}$

$K'_{\mathrm{G}} = \{(\forall X)(peng(X) \rightarrow bird(X)), (\forall X)(bird(X) \wedge peng(X) \rightarrow \neg flies(X))\}$

The ground logical consequences of the revised belief base are:

$$\{bird(tweety), bird(opus), peng(opus), \neg flies(opus)\}$$

In this example we can see that while we have lost some knowledge, *i.e.*, the rule regarding the flying capabilities of birds, we have learned about an exception, *i.e.*, penguins are birds that do not fly and this discovery is precluding us from maintaining a useful general rule. However, we could improve the outcome if we preserve retracted beliefs with a different status according to the mechanism we will present in the next sections.

5 Conditionals and Belief Revision

In this section we will consider different kinds of conditionals and how they can be used in a belief revision system. We will briefly present three main kinds of conditionals, their properties and their use in knowledge representation. Every conditional contains two well distinguished parts: the *antecedent* and the *consequent*. We will study two inference rules on conditionals: modus ponens and antecedent strengthening.

The material conditionals are referred to material implications in most classical logic systems. This type of conditional has the form $\alpha \rightarrow \beta$ and it allows making inferences in different directions. For instance, from α and $\alpha \rightarrow \beta$ we may obtain β applying *modus ponens*. On the other hand, from $\neg\beta$ and $\alpha \rightarrow \beta$ we may infer $\neg\alpha$ applying *modus tolens*. Moreover, the material conditionals satisfy *antecedent strengthening* (If $\alpha \rightarrow \beta$ then $\alpha \wedge \delta \rightarrow \beta$.) This property of material conditionals makes them difficult to use in knowledge representation.

A *counterfactual conditional* is a sentence of the form $\alpha > \beta$ where normally the premise α is either undetermined or rejected (*i.e.*, expected to be false). The counterfactuals have been studied by Lewis [Lew73] and Ginsberg [Gin86]. Each sentence $\alpha > \beta$ is interpreted as "if it were the case that α then β would be the case". The counterfactual conditionals satisfy modus ponens but not antecedent strengthening. If we analyze counterfactuals by means of a truth table we could ensure that, if the antecedent is true then the conditional truth value is equal to the consequent truth value. However, the truth of the counterfactual depends upon more than merely the truth or falsity of the components. Counterfactuals assume the existence of a sphere system centered on a single world i that represents the real world. This is one of most important differences between the model for belief revision proposed by Grove [Gro88] which is a sphere system centered in a set of worlds: the worlds in which the belief set $\mathbf{K}$ holds.

Defeasible conditionals are conditionals in which, if the antecedent is true then "normally" the consequent is true. The sentence $\alpha \succ\!\!- \beta$ is interpreted as "if α holds then normally β holds" or "if α is true then usually β is true". For instance, if $b(X)$ is interpreted as "X is a bird" and $f(X)$ is interpreted as "X flies" then a conditional of the form $b(X) \succ\!\!- f(X)$ is interpreted as: "every individual X that is a bird is *normally* a flying individual". However, if we believe that *Poly* is a bird we cannot conclude that *Poly* flies (*i.e.*, we cannot apply modus ponens in the strong sense). The reason for this is that a defeasible conditional does not allow "skipping" from the antecedent to the consequent since it is a general rule that holds in *normal conditions*, although it is not easy to determine when. Defeasible conditionals do not satisfy antecedent strengthening. That is, from $\alpha \succ\!\!- \delta$ it is not possible to infer $\alpha \wedge \beta \succ\!\!- \delta$. In general, they are used as inference rules instead of language objects, that is, they are on the metalevel. For instance, the *default rules* proposed by Reiter [Rei80] are an example of the use of defeasible conditionals as inference rules. The general form of a default rule is $\frac{\alpha:\beta}{\delta}$ and it is interpreted as: if α is true and β may be consistently assumed then δ is concluded. Among default rules, there is a more specific subclass, called *normal default rules*, and they are of the form $\frac{\alpha:\beta}{\beta}$. A typical example of this kind of rule is:

$$\frac{bird(X) : flies(X)}{flies(X)}$$

In other words, for every individual X which is a bird and it can be consistently assumed that it flies, X is a flying individual.

A formalism in which defeasible conditionals are used as inference rules are the *argumentative systems* [SL92,SCnG94,Pra93,Vre93]. In these systems, each sentence of the form $\alpha \succ\!\!- \beta$ is a tentative inference rule that can be used to obtain new conclusions. Next we will present a revision operator that generates defeasible conditionals from a revision operator upon belief bases represented in a first order language.

5.1 *Generating defeasible conditionals by means of revisions*

Here we will study the generation of defeasible conditionals from a process of belief revision. That idea was formally introduced by Alchourrón [Alc93] upon modal systems and by Falappa and Simari [FS95] upon knowledge based systems.

Suppose that, in a revision process, we eliminate a conditional sentence of the form $\forall(X)(\alpha(X) \rightarrow \beta(X))$. This sentence ensures that any object X satisfying the relation α is an object satisfying the relation β. It can also express that any object satisfying the relation $\neg\beta$ is an object satisfying the relation $\neg\alpha$. If we eliminate such a sentence surely we have received new information inconsistent with it and which is more important. Therefore, one of the following cases may occur:

(1) We have received information regarding some individual satisfying the relation α but not satisfying β.
(2) We have received information regarding some individual satisfying the relation $\neg\beta$ but not satisfying $\neg\alpha$.

In this case, we could discard the refuted rule because we have accepted that it has an exception. We will resume Example 17 in which this policy produces too much loss of information. From the revised set we cannot infer that Tweety flies because we do not have the rule establishing that all birds fly anymore. A way to conclude that Tweety flies could be to use a defeasible conditional (in a disjoint set from the original) determining that, if X is a bird and there is no evidence against the fact that Tweety flies, then we can conclude that Tweety is able to do it. This can be represented by the default rule:

$$\frac{bird(X) : flies(X)}{flies(X)}$$

or by a defeasible rule such as $bird(X) \succ\!\!- flies(X)$ in argumentative systems. Next, we will present a framework to define a revision operator by a set of sentences that generates defeasible conditionals by product of a revision process.

From now on, the epistemic state of an agent will be represented by a tuple of the form $[\![K, \Delta]\!]$ (called *knowledge structure*) where K is a subset of $\mathfrak{L}^+$ and Δ is a set of the form:

$$\Delta = \{\alpha \succ\!\!- \beta : \alpha, \beta \in \mathfrak{L}^+\}$$

Each sentence of K is a well formed formula in $\mathfrak{L}^+$ and it contains those sentences undefeasible in one moment of time. The set K is called *strong* or *undefeasible knowledge* and it is split into two sets K_{P} and K_{G} such that K_{P} represents particular knowledge, K_{G} represents general knowledge, $K = K_{\mathrm{P}} \cup K_{\mathrm{G}}$ and $K_{\mathrm{P}} \cap K_{\mathrm{G}} = \varnothing$. On the other hand, each sentence in Δ is a defeasible conditional representing a tentative inference rule to handle incomplete information. The set Δ is called *defeasible knowledge*. The idea is that, some defeasible rule of the form $\alpha \succ\!\!- \beta$ in Δ is the transformation of some rule $\alpha \rightarrow \beta$ previously included in the strong knowledge but eliminated by some change operator. Instead of fully eliminating that sentence, we propose to preserve a syntactic transformation of it in a different set.

Definition 18 *Let $\delta = (\forall X_1 \ldots X_n)\alpha \rightarrow \beta$ be a material implication in $\mathfrak{L}^+$. A* positive transformation *of δ, noted by $T^+(\delta)$, is a sentence of the form $\alpha \succ\!\!- \beta$; a* negative transformation *of δ, noted by $T^-(\delta)$, is a sentence of the form $\neg\beta \succ\!\!- \neg\alpha$.*

Different from material implications, variables in defeasible conditionals are considered free. The way in which a defeasible conditional of the form $\alpha \succ\!\!- \beta$ is interpreted is: "reasons to believe in the antecedent α provide reasons to believe in the consequent β". Now, we will define a revision operator upon a knowledge structure.

Definition 19 *Let $[\![K, \Delta]\!]$ be a knowledge structure, "$\circ$" an operator of kernel (partial meet) revision by a set of sentences for K and A a set of sentences. The* kernel *(*partial meet*)* composed revision *of $[\![K, \Delta]\!]$ with respect to A is defined as $[\![K, \Delta]\!] \star A = [\![K', \Delta']\!]$ such that $K' = K \circ A$ and $\Delta' = \Delta \cup \Delta'_1 \cup \Delta'_2$ where:*

$$\Delta'_1 = \{\mathbf{true} \succ\!\!- \alpha : \alpha \in (K_{\mathrm{P}} \setminus K \circ A)\}$$

$$\Delta'_2 = \{\, T^+(\alpha) : \alpha \in (K_{\mathrm{G}} \setminus K \circ A)\} \cup \{\, T^-(\alpha) : \alpha \in (K_{\mathrm{G}} \setminus K \circ A)\}$$

The set K' contains the undefeasible revised beliefs, Δ'_1 is the transformation in defeasible rules of particular beliefs (also called assumptions [GSCn98]) eliminated from K whereas Δ'_2 is the transformation of general beliefs eliminated from K into defeasible rules.

6 Argumentative Systems

Here we will introduce argumentative systems as in [SL92,SCnG94], we will define the new epistemic model and we will show some examples of the application of the composed revision.

The derivations in argumentative systems make use of some ground instances (*i.e.*, without free variables) of defeasible rules in Δ. So we will use the set $\Delta^{\downarrow}$ of all ground instances of members of Δ produced by replacing consistently all variables by constants in $\mathfrak{L}^{+}$.

Typically, argumentative systems use sets of undefeasible and defeasible beliefs. In other words, they use a knowledge structure $[\![K, \Delta]\!]$ such that K represents undefeasible beliefs whereas Δ represents defeasible beliefs.

Given a sentence $\alpha \in \mathfrak{L}$ and a set $\Gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ where each α_i is a sentence in K or a member of $\Delta^{\downarrow}$, a meta-meta-relationship "$\mid\!\sim$", called *defeasible consequence* between Γ and α is established as follows [SL92,SCnG94].

Definition 20 *Let $\Gamma \subseteq K \cup \Delta^{\downarrow}$. A ground literal $\alpha \in \mathfrak{L}^{+}$ is a* defeasible consequence *of a set Γ if and only if there is some sequence $\beta_1, \beta_2, \ldots, \beta_m$ such that $\beta_m = \alpha$ and, for any i, $\beta_i \in \Gamma$ or β_i is a direct consequence of the preceding members of the sequence using modus ponens, weak detachment*[4] *or instantiation of a universally quantified sentence or an instance of an axiom in $\mathfrak{L}^{+}$. The notation $\Gamma \mid\!\sim \alpha$ is an abbreviation of α* is a defeasible consequence of Γ.

We can use the notation $\alpha_1, \ldots, \alpha_n \mid\!\sim \alpha$ instead of $\{\alpha_1, \ldots, \alpha_n\} \mid\!\sim \alpha$, or $K \cup T \mid\!\sim \alpha$ making the distinction between defeasible and undefeasible sentences used in the derivation explicit, where $T \subseteq \Delta^{\downarrow}$. The defeasible inference relation allows the definition of a defeasible consequence operator C as $C(\Gamma) = \{\alpha : \Gamma \mid\!\sim \alpha\}$. This operator is nonmonotonic since some derivations can be invalidated on the arrival of new pieces of information. That is so because defeasible rules belonging to each (defeasible) derivation can be invalidated if the undefeasible knowledge is modified.

Different from classical logic systems, the conclusions in argumentative systems are tentative, so we need some selection mechanism or preference criteria among arguments. First, we will introduce the notion of argument in order to define a preference criteria among them.

Definition 21 (Simari & Loui [SL92]) *Given a knowledge structure $[\![K, \Delta]\!]$ we say that a subset T of $\Delta^{\downarrow}$ is an* argument *for a ground literal $\alpha \in \mathfrak{L}^{+}$ in the*

[4] Weak detachment is like modus ponens but using "$\succ\!\!-$" instead of "$\rightarrow$".

context K, noted by $\langle T, \alpha \rangle$, if and only if:

(1) $K \cup T \mathrel{|\!\sim} \alpha$.
(2) $K \cup T \mathrel{|\!\not\sim} \perp$.
(3) There is no $T' \subset T$ *such that* $K \cup T' \mathrel{|\!\sim} \alpha$.

The concept of an argument is similar to the concept of an explanation. However, the first one is defined in terms of a defeasible inference relation and uses context knowledge, whereas the second one is defined in terms of a classical inference relation.

Definition 22 (Simari & Loui [SL92]) *Let $\langle T, \alpha \rangle$ be an argument for α in a context K. We say that $\langle S, \beta \rangle$ is a* subargument *of $\langle T, \alpha \rangle$ if and only if $\langle S, \beta \rangle$ is an argument for β and $S \subseteq T$. This relation is noted as $\langle S, \beta \rangle \subseteq \langle T, \alpha \rangle$, overloading the inclusion symbol upon sets.*

Proposition 23 (Simari & Loui [SL92]) *Every argument $\langle T, \alpha \rangle$ contains the following trivial subarguments: $\langle T, \alpha \rangle$ and $\langle \varnothing, \beta \rangle$ for any $\beta \in Cn(K)$.*

Definition 24 (Simari & Loui [SL92]) *Given two arguments $\langle T_1, \alpha_1 \rangle$ and $\langle T_2, \alpha_2 \rangle$ we say that they are* in disagreement*, noted by $\langle T_1, \alpha_1 \rangle \bowtie \langle T_2, \alpha_2 \rangle$, if and only if $K \cup \{\alpha_1, \alpha_2\} \vdash \perp$.*

Definition 25 [SL92,SCnG94] *An argument $\langle T_1, \alpha_1 \rangle$* counterargues *to $\langle T_2, \alpha_2 \rangle$ at a literal α, noted by $\langle T_1, \alpha_1 \rangle \otimes\!\xrightarrow{\alpha} \langle T_2, \alpha_2 \rangle$, if and only if, there is some subargument $\langle T, \alpha \rangle$ of $\langle T_2, \alpha_2 \rangle$ such that $\langle T_1, \alpha_1 \rangle \bowtie \langle T, \alpha \rangle$.*

Since argument conclusions are tentative we can have situations in which there are arguments with contradictory conclusions. In such case, it is necessary to get a preference criteria. For instance, we will use the criteria of specificity introduced by Poole [Poo85].

Definition 26 (Poole [Poo85], Simari & Loui [SL92]) *Let $\langle T_1, \alpha_1 \rangle$ and $\langle T_2, \alpha_2 \rangle$ be two arguments in the context K. We say that $\langle T_1, \alpha_1 \rangle$ is* strictly more specific *than $\langle T_2, \alpha_2 \rangle$, noted by $\langle T_1, \alpha_1 \rangle \succ_{\text{spec}} \langle T_2, \alpha_2 \rangle$, if and only if:*

(1) For any ground literal $\beta \in \mathfrak{L}^+$ such that $K_{\mathrm{G}} \cup \{\beta\} \cup T_1 \mathrel{|\!\sim} \alpha_1$ and $K_{\mathrm{G}} \cup \{\beta\} \mathrel{|\!\not\sim} \alpha_1$ then $K_{\mathrm{G}} \cup \{\beta\} \cup T_2 \mathrel{|\!\sim} \alpha_2$.
(2) There is a ground literal $\delta \in \mathfrak{L}^+$ such that
  (a) $K_{\mathrm{G}} \cup \{\delta\} \cup T_2 \mathrel{|\!\sim} \alpha_2$ (activates T_2).
  (b) $K_{\mathrm{G}} \cup \{\delta\} \mathrel{|\!\not\sim} \alpha_2$ (nontriviality condition).
  (c) $K_{\mathrm{G}} \cup \{\delta\} \cup T_1 \mathrel{|\!\not\sim} \alpha_1$ (does not activate T_1).

The term activates *is used with the following meaning: together with K_{G} the argument T_i is enough to construct a defeasible derivation of α_j.*

Definition 27 [SCnG94,SL92] *Given two arguments $\langle T_1, \alpha_1\rangle$ and $\langle T_2, \alpha_2\rangle$ we say that $\langle T_1, \alpha_1\rangle$* defeats *$\langle T_2, \alpha_2\rangle$ at a literal α, noted by $\langle T_1, \alpha_1\rangle \gg_{\mathrm{def}} \langle T_2, \alpha_2\rangle$, if and only if there is a subargument $\langle T, \alpha\rangle$ of $\langle T_2, \alpha_2\rangle$ such that:*

Proper Defeater: $\langle T_1, \alpha_1\rangle \succ_{\mathrm{spec}} \langle T, \alpha\rangle$; or

Blocking Defeater: $\langle T_1, \alpha_1\rangle$ is incomparable (by specificity) to $\langle T, \alpha\rangle$.

Example 28 *Given $K = K_{\mathrm{P}} \cup K_{\mathrm{G}}$ where $K_{\mathrm{P}} = \{h(o), p(t)\}$ and $K_{\mathrm{G}} = \{(\forall X)p(X) \rightarrow b(X), (\forall X)h(X) \rightarrow b(X)\}$, and the set of defeasible conditionals:*[5]

$$\Delta = \{b(X) \succ\!\!- f(X), p(X) \wedge b(X) \succ\!\!- \neg f(X), b(X) \wedge f(X) \succ\!\!- w(X)\}$$

We have the following relations between arguments:

- Argument: $\langle\{b(o) \succ\!\!- f(o), b(o) \wedge f(o) \succ\!\!- w(o)\}, w(o)\rangle$.
- Disagreement: $\langle\{p(t) \wedge b(t) \succ\!\!- \neg f(t)\}, \neg f(t)\rangle \bowtie \langle\{b(t) \succ\!\!- f(t)\}, f(t)\rangle$.
- Counterargument:
  $\langle\{p(t) \wedge b(t) \succ\!\!- \neg f(t)\}, \neg f(t)\rangle \otimes\!\xrightarrow{f(t)} \langle\{b(t) \succ\!\!- f(t), b(t) \wedge f(t) \succ\!\!- w(t)\}, w(t)\rangle$.
- More specific: $\langle\{p(t) \wedge b(t) \succ\!\!- \neg f(t)\}, \neg f(t)\rangle \succ_{\mathrm{spec}} \langle\{b(t) \succ\!\!- f(t)\}, f(t)\rangle$.
- Defeat:
  $\langle\{p(t) \wedge b(t) \succ\!\!- \neg f(t)\}, \neg f(t)\rangle \gg_{\mathrm{def}} \langle\{b(t) \succ\!\!- f(t), b(t) \wedge f(t) \succ\!\!- w(t)\}, w(t)\rangle$.

Definition 29 [SCnG94] *Let $\langle T, \alpha\rangle$ be an argument for α in the context K. We say that $\langle T, \alpha\rangle$ is a* justification *for α if for any counterargument $\langle S, \beta\rangle$ of $\langle T, \alpha\rangle$ it holds that $\langle T, \alpha\rangle \gg_{\mathrm{def}} \langle S, \beta\rangle$.*

From now on, the notation $[\![K, \Delta]\!] \vdash \alpha$ is referred to a classical derivation α since it only uses sentences of K. On the other hand, the notation $[\![K, \Delta]\!] \mid\!\sim \alpha$ is referred to a derivation of α using ground instances of sentences in Δ.

Example 30 *In the Example 17 we have the set $K = K_{\mathrm{P}} \cup K_{\mathrm{G}}$ such that:*

$K_{\mathrm{P}} = \{\mathit{bird}(\mathit{tweety}), \mathit{peng}(\mathit{opus})\}$

$K_{\mathrm{G}} = \{(\forall X)(\mathit{peng}(X) \rightarrow \mathit{bird}(X)), (\forall X)(\mathit{bird}(X) \rightarrow \mathit{flies}(X))\}$

Suppose that the knowledge structure is $[\![K, \Delta]\!]$ where $\Delta = \varnothing$. In this case we have that:

$$C([\![K, \Delta]\!]) = \{\mathit{bird}(\mathit{tweety}), \mathit{peng}(\mathit{opus}), \mathit{bird}(\mathit{opus}), \mathit{flies}(\mathit{tweety}), \mathit{flies}(\mathit{opus})\}$$

[5] Literals: o=*opus* and t=*tweety*. Predicates: p=*penguin*, b=*bird*, h=*hawk*, f=*flies* and w=*winged*.

The set of defeasible consequences of $[\![K, \Delta]\!]$ *is equal to the set of logical consequences of* K *since the set of defeasible rules is empty. Then, we receive the explanation for* $\neg \mathit{flies}(\mathit{opus})$*:*

$$A = \{\mathit{bird}(\mathit{opus}), \mathit{peng}(\mathit{opus}), (\forall X)(\mathit{bird}(X) \wedge \mathit{peng}(X) \rightarrow \neg \mathit{flies}(X))\}$$

and we obtain a new belief base $K' = K \circ A = K'_{\mathrm{P}} \cup K'_{\mathrm{G}}$ *such that:*

$K'_{\mathrm{P}} = \{\mathit{bird}(\mathit{tweety}), \mathit{bird}(\mathit{opus}), \mathit{peng}(\mathit{opus})\}$

$K'_{\mathrm{G}} = \{(\forall X)(\mathit{peng}(X) \rightarrow \mathit{bird}(X)), (\forall X)(\mathit{bird}(X) \wedge \mathit{peng}(X) \rightarrow \neg \mathit{flies}(X))\}$

However, the sentences erased are not fully forgotten but they are stored (transformed) as defeasible rules. That is, the new knowledge structure is $[\![K', \Delta']\!] = [\![K, \Delta]\!] \star A$ *where* $K' = K \circ A$ *and:*

$$\Delta' = \{\mathit{bird}(X) \succ\!\!- \mathit{flies}(X), \neg \mathit{flies}(X) \succ\!\!- \neg \mathit{bird}(X)\}$$

Then, we have the following defeasible conclusions:

$C([\![K', \Delta']\!]) = \{\mathit{bird}(\mathit{tweety}), \mathit{bird}(\mathit{opus}), \mathit{peng}(\mathit{opus}), \neg \mathit{flies}(\mathit{opus}), \mathit{flies}(\mathit{tweety})\}$

The last literal is derived using defeasible rules. In other words, using defeasible rules we can extend the conclusions inferred using classical consequence.

6.1 Epistemic Model

Now we will define the new set of epistemic attitudes. Let $[\![K, \Delta]\!]$ be a knowledge structure and α a ground literal in $\mathfrak{L}^+$. The possible epistemic attitudes towards α are:

(1) **Acceptance**: If there is a *justification* $\langle T, \alpha \rangle$.
(2) **Rejection**: If for any possible argument $\langle T, \alpha \rangle$ there is at least an undefeated proper defeater of $\langle T, \alpha \rangle$.
(3) **Indeterminate**: If there is no argument $\langle T, \alpha \rangle$.
(4) **Indefinite**: If for any possible argument $\langle T, \alpha \rangle$ there is no undefeated proper defeater of $\langle T, \alpha \rangle$ but there is at least an undefeated blocking defeater of $\langle T, \alpha \rangle$.

Now, we will give a test for defeasible conditionals assuming the existence of a revision operator upon a knowledge structure and considering defeasible conditionals as a meta-linguistic relation, not as a connective of the object language.

Test for Defeasible Conditionals: $\alpha(X) \succ\!\!- \beta(X) \in [\![K, \Delta]\!] \star A$ if and only if $\alpha(X) \succ\!\!- \beta(X) \in \Delta$ or $(\forall X)(\alpha(X) \rightarrow \beta(X)) \in K$ and $[\![K, \Delta]\!] \star A \vdash \alpha(t) \wedge \neg\beta(t)$ for some ground term t.

Note that this test for defeasible conditionals is formulated in terms of a classical inference relation "$\vdash$". The idea is the following: a defeasible conditional $\alpha \succ\!\!- \beta$ belongs to the revised knowledge structure if and only if it was in the original knowledge structure or $\alpha \rightarrow \beta$ was in K but not in the undefeasible revised knowledge.

Example 31 *Suppose we have the knowledge structure* $[\![K_0, \Delta_0]\!]$ *where:*

$K_0 = \{p(a), p(b), p(c), p(d), q(a), q(b), (\forall X)p(X) \rightarrow s(X), (\forall X)q(X) \rightarrow t(X)\}$

$\Delta_0 = \varnothing$

The ground sentences inferred from this belief base are:

$$\{p(a), p(b), p(c), p(d), q(a), q(b), s(a), s(b), s(c), s(d), t(a), t(b)\}$$

At some given instant, we receive the following explanation A_0 *for* $\neg s(a)$*:*

$$\{p(a), q(a), (\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X)\}$$

If we are making a kernel revision by a set of sentences we must give up the minimally inconsistent subsets of $K_0 \cup A_0$*. That is, we must cut every set in* $(K_0 \cup A_0)^{\perp\!\!\perp}\bot = \{H_1, H_2\}$ *where:*

$$H_1 = \{p(a), q(a), (\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X), (\forall X)p(X) \rightarrow s(X)\}$$

$$H_2 = \{p(b), q(b), (\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X), (\forall X)p(X) \rightarrow s(X)\}$$

We can eliminate one of the common rules to cut both minimally entailment sets. Suppose that the belief $(\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X)$ *is better or more plausible than* $(\forall X)p(X) \rightarrow s(X)$*. Then, the new knowledge structure is* $[\![K_1, \Delta_1]\!] = [\![K_0, \Delta_0]\!] \star A_0$ *where:*

$K_1 = \{p(a), p(b), p(c), p(d), q(a), q(b), (\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X), (\forall X)q(X) \rightarrow t(X)\}$

$\Delta_1 = \{p(X) \succ\!\!- s(X), \neg s(X) \succ\!\!- \neg p(X)\}$

Note that in this knowledge structure we can infer the ground sentences:

$$\{p(a), p(b), p(c), p(d), q(a), q(b), \neg s(a), \neg s(b), t(a), t(b)\}$$

However, we believe in $p(c)$ *and* $p(d)$ *but we cannot conclude classically* $s(c)$

and $s(d)$ although these sentences are consistent with K_1. If we use the defeasible rule $p(X) \succ\!\!- s(X)$ we can see that there are justifications for $s(c)$ and $s(d)$. Therefore, $s(c)$ and $s(d)$ will be accepted in the revised knowledge structure.

Example 32 *Consider the knowledge structure produced in the above example. At some point, we receive the following explanation A_1 for $s(a)$:*

$$\{p(a), q(a), u(a), (\forall X)(p(X) \wedge q(X) \wedge u(X)) \rightarrow s(X)\}$$

Now, we must eliminate the minimally inconsistent sets of $K_1 \cup A_1$. The one set in this condition is:

$$\{p(a), q(a), u(a), (\forall X)(p(X) \wedge q(X) \wedge u(X)) \rightarrow s(X), (\forall X)(p(X) \wedge q(X)) \rightarrow \neg s(X)\}$$

If the new rule is considered better than the sentence in K_1 then the revised knowledge structure is: $[\![K_2, \Delta_2]\!] = [\![K_1, \Delta_1]\!] \star A_1$ where:

$$\begin{aligned} K_2 = {} & \{p(a), p(b), p(c), p(d), q(a), q(b), u(a)\} \cup \\ & \{(\forall X)(p(X) \wedge q(X) \wedge u(X)) \rightarrow s(X), (\forall X) q(X) \rightarrow t(X)\} \\ \Delta_2 = {} & \{p(X) \succ\!\!- s(X), \neg s(X) \succ\!\!- \neg p(X)\} \cup \\ & \{(p(X) \wedge q(X)) \succ\!\!- \neg s(X), s(x) \succ\!\!- (\neg p(X) \vee \neg q(X))\} \end{aligned}$$

In this knowledge structure we can infer the following ground sentences:

$$\{p(a), p(b), p(c), p(d), q(a), q(b), u(a), s(a), t(a), t(b)\}$$

The defeasible ground consequences of this set are: $\{s(b), s(c), s(d), \neg s(b)\}$. Again, we have extended the set of ground conclusions. But in this case we can (defeasibily) infer either $s(b)$ or $\neg s(b)$ since we can construct the arguments $\langle \{p(b) \succ\!\!- s(b)\}, s(b) \rangle$ and $\langle \{(p(b) \wedge q(b)) \succ\!\!- \neg s(b)\}, \neg s(b) \rangle$. Since the argumentative systems can treat this kind of contradictions (by means of a preference relation between arguments) it could be the case that one argument defeats another. Using the specificity *criteria proposed in [Poo85,SL92] we could conclude that the argument for $\neg s(b)$ defeats the argument for $s(b)$. Therefore, $\neg s(b)$ will be accepted in the knowledge structure.*

It is easy to check that the test for defeasible conditionals is satisfied in both examples.

6.2 Which beliefs should be revised?

A revision operator can modify either the undefeasible or the defeasible knowledge. The main problem is to determine when some piece of information is undefeasible or defeasible. A simple solution could be to incorporate knowledge directly upon the defeasible knowledge. But this solution is too simple and it is not very realistic. In our perspective, the qualification of the knowledge is dynamic, that is, it evolves with time and the incorporation of new information. When an agent incorporates knowledge it typically incorporates it into its undefeasible knowledge. But, should it be possible to consider this new knowledge as defeasible knowledge if it were not actually so? Our position is that the knowledge is undefeasible until we discover new information inconsistent with it. That is, suppose we believe undefeasibly that all private enterprizes give an optimal service ($\alpha = (\forall X)(\mathit{enterprise}(X) \wedge \mathit{private}(X)) \rightarrow \mathit{good\ service}(X))$. Then we receive new information saying that private enterprizes with foreign capitals provide a bad service ($\beta = (\forall X)(\mathit{enterprise}(X) \wedge \mathit{private}(X) \wedge \mathit{foreigner}(X)) \rightarrow \neg \mathit{good\ service}(X))$. And we receive new information saying that *unicom* is a phone company with foreign capitals, its rate is very high and the service is not very good. At this moment we change the status of the belief. In other words, we will believe undefeasibly in each belief until we note new and more plausible information in contradiction with them. In this case, we do not undefeasibly believe in α. We will undefeasibly believe in β (since it is more specific than α) and α could be considered as a tentative belief. That is, for every private enterprize with foreign capitals we will believe that it does not provide a good service. On the other hand, for those private enterprizes with unknown source of capital we could believe that they give a good service.

This intuition is modelled with the revision operator by a set of sentences upon defeasible systems. Suppose that $[\![K, \Delta]\!]$ represents the epistemic state of an agent. After the revision with respect to some explanans A we modify the undefeasible knowledge by a non-prioritized revision operator. However, those sentences eliminated in the revision process are not fully discarded, but stored as tentative rules (with a different status). This mechanism has two advantages:

(1) *Dynamic classification of beliefs*: classify beliefs dynamically as undefeasible or defeasible.
(2) *Minimal change*: preserve as much old information as possible.

The idea of *minimal change* is one of the main principles of theory change. The concept of *dynamic classification* has been frequently used in the evolution of humanity's knowledge. For instance, the belief establishing that all metals are solid under normal conditions of temperature and pressure was undefeasible for years (maybe centuries). However, at some point, it was discovered that mercury is a metal in liquid state under said conditions. Precisely at this moment the status of the belief about the solidity property of metals was modified. By analogy, at some moment we believed that the freezing temperature of water was 0° Celsius. This belief was undefeasible until it was discovered that this property holds only under normal conditions of pressure. If the pressure is high, the freezing point could be a lower temperature. Therefore, a more specific rule was found for determining the exact freezing point of water. On the other hand, the belief saying that the freezing temperature of water is 0^o Celsius could be viewed as a default rule that holds within “normal” worlds.

At the base of the above reasoning, we think that beliefs (rules, facts, defaults, arguments) are dynamically classified as undefeasible or defeasible by successive revisions. Someone could believe that it would be simpler to incorporate beliefs directly upon the defeasible knowledge. However, this policy could decrease the inference power of an agent. Consider the following example. Think of a rule of the kind *is-a* (very often used in database relationships), for instance $(\forall X) argentinian(X) \rightarrow south\ american(X)$. This rule represents an is-a relationship: every Argentinean is a South American. Moreover, if we know that John is not a South American we can conclude that John is not an Argentinean. This conclusion is not possible if the above rule were defeasible.

We could think that every is-a rule is an undefeasible one. However, is-a rules are not the only undefeasible rules. Many prototypical properties of objects are undefeasible too. For instance, every man has a heart as vital organ. This rule is not an is-a rule and it makes reference to properties of the object man. Therefore, we are giving one more argument to treat knowledge as undefeasible until we discover new and better information.

7 Conclusions

We have presented a new kind of non-prioritized revision operator based on the use of explanations. The idea is that an agent, before incorporating information which is inconsistent with its knowledge, requests an explanation supporting it. We distinguish two parts in every explanation: an explanans, represented by set of sentences supporting some belief, and an explanandum, which is the final conclusion. We present a deductive notion of explanation, giving some postulates for it.

We propose that every explanation contains rules and particular knowledge. If the sentences in the explanans are better or more plausible than the sentences in the original belief base, then the explanation is incorporated. We have defined two kinds of revision operator: kernel and partial meet revision by a set of sentences. These operators may partially accept the new information and we give representation theorems for them.

Finally, we presented a framework oriented to defeasible reasoning. We showed how defeasible conditionals can be generated using the revision operator by a set of sentences upon knowledge structures. This approach is sound because it preserves consistency in the undefeasible knowledge and it provides a mechanism to dynamically qualify the beliefs as undefeasible or defeasible. Moreover, it provides a more complete set of epistemic attitudes and extends the inference power of knowledge based systems.

Acknowledgements

We would like to thank David Makinson, Sven Ove Hansson and Eduardo Fermé for helpful discussions on some of the concepts included in this paper.

References

[AGM85] Carlos Alchourrón, Peter Gärdenfors, and David Makinson. *On the Logic of Theory Change: Partial Meet Contraction and Revision Functions*. *The Journal of Symbolic Logic*, 50:510–530, 1985.

[Alc93] Carlos Alchourrón. *Defeasible Conditionals as General Conditionals Plus Revision Theory*. Technical report, Universidad de Buenos Aires, Argentina, 1993.

[BB93] Craig Boutilier and Verónica Becher. *Abduction as Belief Revision: A Model of Preferred Explanations*. Technical report, University of British Columbia, Department of Computer Science, Vancouver, British Columbia, Canadá, 1993.

[FH98] Eduardo L. Fermé and Sven Ove Hansson. *Selective Revision*. *Studia Logica*, in press, 1998.

[FS95] Marcelo A. Falappa and Guillermo R. Simari. *Condicionales Derrotables a partir de la Revisión de Condicionales Estrictos*. *Jornadas Argentinas de Informática e Investigación Operativa*, JAIIO '95:7.13–7.28, 1995.

[Fuh97] André Fuhrmann. *An Essay on Contraction*. Studies in Logic, Language and Information, CSLI Publications, Stanford, California, 1997.

[Gä88] Peter Gärdenfors. *Knowledge in Flux: Modelling the Dynamics of Epistemic States*. The MIT Press, Bradford Books, Cambridge, Massachusetts, 1988.

[Gin86] Matthew L. Ginsberg. *Counterfactuals*. *Artificial Intelligence*, 30:35–79, 1986.

[Gro88] Adam Grove. *Two Modellings for Theory Change*. *The Journal of Philosophical Logic*, 17:157–170, 1988.

[GSCn98] Alejandro J. García, Guillermo R. Simari, and Carlos I. Chesñevar. *An Argumentative Framework for Reasoning with Inconsistent and Incomplete Information*. *Workshop on Practical Reasoning and Rationality, Brighton, England*, ECAI '98:13–19, 1998.

[Han91] Sven Ove Hansson. *Belief Contraction Without Recovery*. *Studia Logica*, 50:251–260, 1991.

[Han93] Sven Ove Hansson. *Kernel Contraction*. *The Journal of Symbolic Logic*, 1993.

[Han96] Sven Ove Hansson. *A Textbook of Belief Dymanics: Theory Change and Database Updating*. Uppsala University, Department of Philosophy, Uppsala, Sweden, 1996.

[Han97] Sven Ove Hansson. *Semi-Revision*. *Journal of Applied Non-Classical Logic*, in press, 1997.

[HFCF01] Sven Ove Hansson, Eduardo Fermé, John Cantwell, and Marcelo Falappa. *Credibility Limited Revision*. *The Journal of Symbolic Logic*, 66(4):1581–1596, 2001.

[Lew73] David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973.

[Pag96] Maurice Pagnucco. *The Role of Abductive Reasoning within the Process of Belief Revision*. PhD thesis, Basser Department of Computer Science, University of Sydney, Sydney, Australia, 1996.

[PNF94] Maurice Pagnucco, Abhaya Nayak, and Norman Foo. *Abduction Expansion: Abductive Inference and the Process of Belief Change*. In *Proceedings of the Seventh Australian Conference on Artificial Intelligence*, Armidale, Australia, 1994.

[Poo85] David Poole. *On the Comparison of Theories: Preferring the Most Specific Explanation*. In *Proceedings of IJCAI-85, Los Angeles (CA)*, pages 144–147, 1985.

[Pra93] Hendrik Prakken. *Logical Tools for Modelling Legal Argument*. PhD thesis, Vrije Universiteit, Amsterdan, 1993.

[Rei80] Raymond Reiter. *A Logic for Default Reasoning*. *Artificial Intelligence*, 13:81–132, 1980.

[Res70] Nicholas Rescher. *Scientific Explanation*. The Free Press, New York, 1970.

[SCnG94] Guillermo R. Simari, Carlos I. Chesñevar, and Alejandro J. García. *The Role of Dialectics in Defeasible Argumentation*. *Conferencia Internacional de la Sociedad Chilena de Computación*, pages 111–121, 1994.

[SL92] Guillermo R. Simari and Ron P. Loui. *A Mathematical Treatment of Defeasible Reasoning and its Implementation*. *Artificial Intelligence*, 53:125–157, 1992.

[SS96] Leon Sterling and Ehud Shapiro. *The Art of Prolog*. The MIT Press, 1996.

[Vre93] Gerardus A.W. Vreeswijk. *Studies in Defeasible Argumentation*. PhD thesis, Vrije Universiteit, Amsterdan, 1993.

[WPFS95] Mary-Anne Williams, Maurice Pagnucco, Norman Foo, and Brailey Sims. *Determining Explanations using Transmutations*. In *Proceedings of IJCAI-95*, pages 822–830, 1995.