University of Alberta

# Practical Belief Change

by

Aditya Kumar Ghose ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Computing Science

Edmonton, Alberta
Fall 1995

ISBN 0-612-06213-9

University of Alberta

Library Release Form

**Name of Author:** Aditya Kumar Ghose

**Title of Thesis:** Practical Belief Change

**Degree:** Doctor of Philosophy

**Year this Degree Granted:** 1995



. . . . . . . . . . . . . . . . . . . . . .

Aditya Kumar Ghose
c/o Prof. A. K. Ghose
Flat No. 7SE
64, Lake Road
Calcutta 700 029
India

**Date:** Oct. 6, 1995

**University of Alberta**

**Faculty of Graduate Studies and Research**

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Practical Belief Change** submitted by Aditya Kumar Ghose in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

. . . . . . . . . . . . . . . . . . . .
R. Goebel

. . . . . . . . . . . . . . . . . . . .
N. Foo

. . Bernard Linsky . . .
B. Linsky

. . . . . . . . . . . . . . . . . . . .
R. Elio

. . . . . . . . . . . . . . . . . . . .
J. H. You

. . . . . . . . . . . . . . . . . . . .
F. J. Pelletier

**Date:** Oct. 5, 1995

*To the memory of my grandparents*
*Hrishikesh Chanda*
*Falguni Chanda*
*and*
*Umarani Ghose*

# Abstract

In this study, we seek to provide a framework for the design of practical systems for belief change. We do this through the following steps:

*Competence:* The work of Alchourrón, Gärdenfors and Makinson provides a comprehensive and widely accepted competence theory for the process of belief change. We identify the following major drawbacks in this theory:

- It provides an inadequate account of the process of retracting a belief. Thus, the addition of a belief is duly recorded in the belief state of an agent, but the retraction of a belief is never recorded. This can unduly restrict the space of candidate outcomes of a belief change operation.
- The theory provides no prescription on how beliefs must change when the belief input is not fully credible. Any approach to handling uncertain, or less credible, belief inputs should involve a generalization of techniques applied when the belief inputs are fully credible, instead of requiring a totally distinct set of techniques.
- It is generally agreed that the *principle of informational economy* should guide any strategy for belief change. This requires that beliefs should be discarded as little as possible while effecting belief change. The competence theory of Alchourrón, Gärdenfors and Makinson seeks to satisfy this requirement, but with limited success. As a consequence of the belief representation scheme and an unduly narrow definition of what constitutes success for a belief change operation, beliefs may be unduly discarded by

operators defined within this framework.

- The theory does not specify belief change beyond a single step. Several authors have sought to address this question, but their solutions suffer from the previous three problems.

We develop a theory that accounts for each of the problems mentioned above, and argue that it provides an adequate set of benchmark tests, as well as a suitable starting point for implemented belief change systems.

*Performance:* We present the design of two belief change systems which use a variant of default logic as the belief representation language. The design of the first system preceded the development of the our competence theory and provided the motivation for this theory, by identifying several of the lacunae in the existing definition of competence. The second system was developed using our competence theory as the starting point. These two designs serve to demonstrate that practically implementable systems that satisfy the requirements identified in our competence theory are indeed possible. The use of a default logic variant has several other practical benefits as well, such as the ability to incorporate lazy evaluation strategies in computing belief change.

*Implementation:* Belief change is a computationally hard problem , including our formulation of the problem in the two systems mentioned above. Nevertheless, practical constraints often require tractable solutions, or procedures that exhibit resource-bounded rationality . We present a toolkit of two approaches to address such concerns. First, we define a mapping from the problem of default inference to partial constraint satisfaction problems . The mapping enables us to apply techniques from the area of partial constraint satisfaction to improve the efficiency of procedures for computing default extensions, and hence for computing belief change. Next, we present a set of strategies for computing meaningful partial results in resource-bounded situations, by defining *anytime* procedures for default inference. While much remains to be done in this area, we believe these strategies can provide the basis for fielded applications of problem solvers with a significant belief change component.

# Acknowledgements

My heartfelt thanks to my Ph.D. supervisor Randy Goebel for his guidance, support and friendship. I doubt if there can be a supervisor better than him. Many thanks to the members of my examining committee, Renee Elio, Jeff Pelletier, Bernard Linsky and Jia You for their technical input, as well as their encouragement. Thanks also to Li Yan Yuan for serving on my supervisory committee. I acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and the Institute for Robotics and Intelligent Systems of the Canadian Federal Networks of Centers of Excellence.

Finally, thanks to my friends and colleagues, too numerous to mention, here in Canada, and in India, the U.S., Australia, Japan, Taiwan and elsewhere.

# Contents

# List of Figures

# Chapter 1

# Introduction

Consider the following scenarios:

- A module in a large software system is modified to reflect changing system requirements. Typically, the change is not restricted to the module directly affected, but must propagate throughout the entire system. Changes often violate global consistency requirements and the whole system must undergo some variety of modification to restore consistency. Typically, too, several alternative sets of modifications may be used to achieve the same end result of restoring system consistency.

- A database is updated with new data. The new data together with the existing contents of the database might violate the integrity constraints that the database must satisfy. To restore the database to a state where the integrity constraints are satisfied, some of the existing data may have to be deleted, or the new data may be rejected. Often, several alternative deletions may be used to restore database consistency.

- A robotic agent operating in a dynamic environment is presented with a stream of information from its sensors. New information may often contradict the agents prior beliefs. Rational agents usually require a consistent set of beliefs to reason with, or act on. Since the new beliefs introduce inconsistency into the agent's set of beliefs, the agent must perform some variety of adjustment to restore consistency.

A common thread runs through each of these scenarios. A body of information, which must at all times satisfy a set of consistency requirements, is updated with new information. The new information causes the consistency requirements to be violated. To restore consistency, adjustments need to be made. In making these adjustments, several options exist and a choice has to be made from amongst these alternatives.

We shall refer to this as the problem of *belief change*. The term derives from studies in the areas of intelligent systems, philosophy and cognitive science on models of rational change of beliefs by intelligent agents. In this dissertation, we shall study the problem of belief change using a formulation which is commonly used in each of these areas of inquiry. Under this formulation, beliefs are represented in a formal language, typically the language of formal logic. The applicable consistency requirement is that of logical consistency. Consider an agent which believes that $a$ is true and $a \rightarrow b$ is true. We do not commit to any detailed structure of the agent's belief state at this point but merely require that the beliefs be represented in a logical language and that the agent hold a logically consistent set of beliefs. Let the agent now learn that $\neg b$ is true. Adding this new belief to the current set of beliefs would introduce inconsistency into the belief set, since $a$ and $a \rightarrow b$ together entail $b$. At this level of analysis, one can identify three possible changes that can be made to maintain a consistent set of beliefs. First, belief in $a$ may be discarded. This would generate a consistent set of beliefs consisting of $a \rightarrow b$ and $\neg b$. Second, belief in $a \rightarrow b$ may be discarded, generating a consistent set of beliefs consisting of $a$ and $\neg b$ as an outcome. Third, the new belief may be rejected, resulting in the original consistent set of beliefs being retained. A belief change operation consists of the following three steps:

1. Generating candidate consistent sets of beliefs, given a prior belief state and the belief input.

2. Selecting one of these candidate sets of beliefs to reason with, or act on.

3. Generating a new belief state.

It may appear that that the last two steps are identical, and this has been the approach taken in several existing frameworks, but we shall show in later chapters that this need not be the case.

The process of belief change is ubiquitous in information processing. In addition to the situations described earlier in this section, belief change forms a fundamental component of problem solving in domains as diverse as induction, combining knowledge bases, reasoning about action and planning, to name but a few.

## 1.1 Practical belief change

Much of the existing work on formally characterizing the process of belief change has involved frameworks which require unrealistic assumptions, such as the requirement for identifying maximal subsets of infinitely large, deductively closed sets of sentences that satisfy certain constraints [1], [19], the requirement that it should be possible to prioritize every belief held by an agent with respect to every other [1], [19], [39], or the potentially myopic focus on formalizing belief change as a single-step process, as opposed to a process that is repeated over time [1], [19], [40]. As well, these frameworks are unable to adequately account for contractions (in the sense that the effects of a contraction step do not persist beyond a single step) and cannot handle belief inputs which are uncertain, or less than fully credible. Finally, the currently popular definition of competence in this area, as embodied in the rationality postulates proposed by Alchourrón, Gärdenfors and Makinson [1], [19], fails to provide an adequate specification of the ideal case, and thus has limited use both as a starting point for designing new systems and as a benchmark test for existing systems.

In this dissertation, we seek to provide a framework for the design of practical systems for belief change. We do this through the following steps:

*Competence:* The work of Alchourrón, Gärdenfors and Makinson [1], [19], provides a comprehensive and widely accepted competence theory for the process of belief change. We identify the following major drawbacks in this theory:

- It provides an inadequate account of the process of retracting a belief. Thus, the addition of a belief is duly recorded in the belief state of an agent, but the retraction of a belief is never recorded. This can unduly restrict the space of candidate outcomes of a belief change operation.
- The theory provides no prescription on how beliefs must change when the belief input is not fully credible. Any approach to handling uncertain, or less credible, belief inputs should involve a generalization of techniques applied when the belief inputs are fully credible, instead of requiring a totally distinct set of techniques.
- It is generally agreed that the *principle of informational economy* should guide any strategy for belief change. This requires that beliefs should be discarded as little as possible while effecting belief change. The competence theory of Alchourrón, Gärdenfors and Makinson seeks to satisfy this requirement, but with limited success. As a consequence of the belief representation scheme and an unduly narrow definition of what constitutes success for a belief change operation, beliefs may be unduly discarded by operators defined within this framework.
- The theory does not specify belief change beyond a single step. Several authors, such as [38], [37] and [10] have sought to address this question, but their solutions suffer from the previous three problems.

We develop a theory that accounts for each of the problems mentioned above, and argue that it provides an adequate set of benchmark tests, as well as a suitable starting point for implemented belief change systems.

*Performance:* We present the design of two belief change systems which use a variant of default logic [14] as the belief representation language. The design of the first system preceded the development of the our competence theory and provided the motivation for this theory, by identifying several of the lacunae in the existing definition of competence. The second system was developed using our competence theory as the starting point. These two designs serve to demonstrate that practically implementable systems that satisfy the requirements identified in our competence theory are indeed possible. The use of a default logic variant has several other practical benefits as well, such as the ability to incorporate lazy evaluation strategies in computing belief change.

*Implementation:* Belief change is a computationally hard problem [40], including our formulation of the problem in the two systems mentioned above. Nevertheless, practical constraints often require tractable solutions, or procedures that exhibit resource-bounded rationality [48]. We present a toolkit of two approaches to address such concerns. First, we define a mapping from the problem of default inference to partial constraint satisfaction problems [18]. The mapping enables us to apply techniques from the area of partial constraint satisfaction to improve the efficiency of procedures for computing default extensions, and hence for computing belief change. Next, we present a set of strategies for computing meaningful partial results in resource-bounded situations, by defining *anytime* procedures for default inference. While much remains to be done in this area, we believe these strategies can provide the basis for fielded applications of problem solvers with a significant belief change component.

## 1.2 Outline of presentation

Chapter 2 surveys some of the existing frameworks for belief change. Chapter 3 presents a new competence theory for belief change. Chapter 4 describes two belief change systems based on a variant of default logic. Chapter 5 describes implementation strategies for belief change systems. Chapter 6 summarizes the contributions of this study and outlines possibilities for future work.

# Chapter 2

# Formal Approaches to Belief Change

## 2.1 The AGM Framework

The systematic study of the dynamics of belief change undertaken by Alchourrón, Gärdenfors and Makinson [1], [21], [34], [19] is perhaps the most influential body of work in this area; we shall refer to their formalization as the *AGM framework*. The AGM framework consists of the following components:

1. A scheme for representing the belief state (or *epistemic state*) of an agent.

2. A specification of the kinds of beliefs that are expressible, i.e., the *epistemic attitudes* of an agent.

3. A specification of the kinds of inputs that may drive belief change, i.e., the *epistemic inputs*, and hence a repertoire of possible belief change operations.

4. A set of rationality conditions that constrain the space of allowable belief change operators.

5. A set of constructions of belief change operators that satisfy these rationality conditions.

In the AGM framework, the epistemic state of an agent is represented by a deductively closed propositional theory, sometimes referred to as a *knowledge set*. Given an agent with an epistemic state denoted by $K$, three possible epistemic attitudes are possible with respect to a belief $x$:

1. $x$ is *accepted*. In this case, $K \models x$.

2. $x$ is *rejected*. In this case, $K \models \neg x$.

3. $x$ is *undetermined*. In this case, neither $K \models x$ nor $K \models \neg x$.

Two kinds of epistemic inputs are permitted:

1. *Addition*, where a new belief is incorporated into the existing set of beliefs.

2. *Abrogation*, where a belief is given up from the existing set of beliefs.

This leads to a repertoire of three belief change operations:

1. *Expansion:* A new belief $x$ is added to the current knowledge set $K$, with the guarantee that $K \cup \{x\}$ is satisfiable. The outcome is denoted by $K_x^+$, where, necessarily, $K_x^+ \models x$.

2. *Contraction:* A belief $x$ is given up from the current knowledge set $K$. This operation maps the knowledge set $K$, where, potentially, $K \models x$, to an outcome, denoted by $K_x^-$, where, necessarily, $K_x^- \not\models x$.

3. *Revision:* A new belief $x$ is added to the current knowledge set $K$, where, potentially, $K \cup \{x\}$ is not satisfiable. The outcome is denoted by $K_x^*$, where, necessarily, $K_x^* \models x$.

Expansion involves the straightforward set-theoretic addition of the new belief to the existing knowledge set, and is guaranteed to have a unique outcome. Thus $K_x^+ = Cn(K \cup x)$. Both contraction and revision are non-trivial, and may potentially have multiple candidate outcomes. The operations of contraction and revision can be defined in terms of each other, as shown by the *Levi identity* [32] below:

$$K^*_A = (K^-_{\neg A})^+_A$$

The *Harper identity* [28] ($K^-_A = K^*_{\neg A} \cap K$) similarly defines contraction in terms of revision.

The AGM framework presents a set of rationality postulates for each of the three operations which constrain the space of possible outcomes of these operations. Since expansion is trivial and is guaranteed to have a unique outcome, we shall not present the rationality postulates for expansion here. With contraction and revision, the focus is on enforcing minimal change. This stems from the so-called *principle of informational economy* which requires that as few beliefs be discarded as possible during a belief change operation. In addition, these postulates require that beliefs be represented in the same form before and after a belief change step, that belief change steps succeed, that they be independent of the syntactic form of beliefs, and that the process be reversible.

We begin with the postulates for contraction.

*(1-)* For any sentence $A$ and any knowledge set $K$, $K^-_A$ is a belief set.

*(2-)* $K^-_A \subseteq K$.

*(3-)* If $A \notin K$, then $K^-_A = K$.

*(4-)* If $\not\models A$, then $A \notin K^-_A$.

*(5-)* If $A \in K$, then $K \subseteq (K^-_A)^+_A$.

*(6-)* If $\models A \leftrightarrow B$, then $K^-_A = K^-_B$.

*(7-)* $K^-_A \cap K^-_B \subseteq K^-_{A \wedge B}$.

*(8-)* If $A \notin K^-_{A \wedge B}$, then $K^-_{A \wedge B} \subseteq K^-_A$.

Postulate (1−) requires that beliefs be represented in the same form before and after a belief change step. (2−) requires that no new beliefs be held as a result of a contraction. (3−) requires that if the belief to be contracted is not held, then no change should be made. (4−) requires that every contraction operation succeed, unless the belief being contracted is a logical truth. (5−) is the principle of recovery, which requires that if a belief held in a given belief state is retracted and then added back to the belief state, the outcome contains the initial belief state, i.e., the initial belief state is recovered. (6−) is the principle of irrelevance of syntax, which requires that the outcome of a contraction operation be independent of the syntactic form of the beliefs being contracted. (7−) requires that the retraction of a conjunction of beliefs should not retire any beliefs that are common to the retraction of the same belief set with each individual conjunct. (8−) requires that, when retracting the conjunct of two beliefs $A$ and $B$ forces us to give up $A$, then in retracting $A$, we do not give up any more than in retracting the conjunction of $A$ and $B$.

The postulates for revision are as follows.

*(1*)* For any sentence $A$ and any knowledge set $K$, $K^*_A$ is a knowledge set.

*(2*)* $A \in K^*_A$.

*(3*)* $K^*_A \subseteq K^+_A$.

*(4*)* If $\neg A \notin K$, then $K^+_A \subseteq K^*_A$.

*(5*)* $K^*_A = \perp$ iff $\models \neg A$.

*(6*)* If $\models A \leftrightarrow B$, then $K^*_A = K^*_B$.

*(7*)* $K^*_{A \wedge B} \subseteq (K^*_A)^+_B$.

*(8*)* If $\neg B \notin K^*_A$, then $(K^*_A)^+_B \subseteq K^*_{A \wedge B}$.

As before, *(1*)* requires that belief states be represented in the same form before and after a revision operation, while *(2*)* requires that the revision operation succeed. *(3*)* requires that the revision of a knowledge set with a belief be contained in the expansion of the knowledge set with the same belief. In the case that the belief is inconsistent with the knowledge set, this is trivially true since the expansion is a contradiction. In the case that the new belief is consistent with the knowledge set, *(4*)* requires that the expansion be contained in the revision, which, given *(3*)* implies that the revision equal the expansion. *(5*)* requires that the revision of a knowledge set with a new belief be a contradiction if and only if the new belief is a contradiction. As with contractions, *(6*)* is the principle of irrelevance of syntax, which requires that the outcome of a revision be independent of the syntactic form of the epistemic inputs. *(7*)* and *(8*)* are generalizations of *(3*)* and *(4*)* to the case of iterated revisions. Thus *(7*)* requires that the minimal change required to incorporate both $A$ and $B$ in $K$ should be contained in the expansion with $B$ of the revision of $K$ with $A$. In the case that $B$ is consistent with the revision of $K$ with $A$, *(8*)* requires that the expansion be contained in the revision, which together with *(7*)*, implies that the revision of $K$ with $A \wedge B$ be equal to the expansion with $B$ of the revision of $K$ with $A$.

In the AGM framework, constructions for revision operators are defined via the Levi identity, i.e., as contractions followed by expansions. A revision operator defined using a contraction operator via the Levi identity satisfies (1$*$) through (8$*$) if the contraction operator satisfies (1$-$) through (8$-$). We shall focus on presenting the AGM constructions for contraction operators. Contraction of a belief $x$ begins with the *removal* of that belief from the current belief state. Let the *removal* of $x$ from $A$, denoted by $A \downarrow x$, be defined as:

$$A \downarrow x = \{B \subseteq A \mid B \not\models x, \forall C : B \subset C \subseteq A \Rightarrow C \models x\}$$

In general, the removal operation may generate multiple candidate belief states, on which a further operation must be performed to generate one unique resultant belief state. The AGM framework defines a toolkit of four approaches to this operation. The first, called *partial meet contraction*, involves the application of a *selection function*. Let $S$ be a selection function that selects a nonempty subset of $K \downarrow x$ (provided $K \downarrow x$ is nonempty, $\emptyset$ otherwise). Essentially, $S$ picks those subsets in $K \downarrow x$ that are epistemologically most entrenched. The partial meet contraction operator $-$, uses this function to generate the final outcome in the following manner: $K_x^- = \bigcap S(K \downarrow \neg x)$

(the contraction step fails if $x$ is logically valid). Let $M(K)$ stand for the family of all the sets $K \downarrow x$, where $x$ is any proposition in $K$ that is not logically valid. Let $\leq$ be a relation defined on $M(K)$. Let:

$$S(K \downarrow x) = \{K' \in K \downarrow x \mid K'' \leq K' \text{ for all } K'' \in K \downarrow x\}$$

Any partial meet contraction operator for which the selection function $S$ is defined in this manner, and for which the relation $\leq$ is transitive, satisfies all the AGM postulates for contraction [19]. The second approach involves taking the intersection of all the candidate outcomes generated by the removal operation; this is called *full meet contraction*. A revision operator based on full meet contraction discards all beliefs but the consequences of the new belief; this approach is therefore unintuitive. The third approach, *maxichoice contraction*, uses a different variety of selection function to pick exactly one of the candidate outcomes. This approach has undesirable consequences as well. A revision operator based on maxichoice contraction results in knowledge sets which are *complete*, i.e., for any belief $y$, the knowledge set necessarily commits to either $y$ or $\neg y$. The fourth approach involves using a specially constrained class of total orderings (called *epistemic entrenchment*), defined on the entire language, to decide which beliefs to retain and which to discard. Let $x \preceq_K y$, where $\preceq_K$ is the epistemic entrenchment relation associated with the belief set $K$, denote that $x$ is at most as entrenched as $y$. The relation $\preceq_K$ must satisfy the following conditions:

*(EE1)* If $x \preceq_K y$ and $y \preceq_K z$, then $x \preceq_K z$.

*(EE2)* If $x \models y$ then $x \preceq_K y$.

*(EE3)* For any $x$ and $y$, $x \preceq_K x \wedge y$ or $y \preceq_K x \wedge y$.

*(EE4)* When $K \neq K_\perp$, $x \notin K$ iff $x \preceq_K y$ for all $y$.

*(EE5)* If $y \preceq_K x$ for all $y$, then $\models x$.

*(EE1)* requires that the $\preceq_K$ relation be transitive. *(EE2)* is motivated by the criterion for informational economy. If $x$ logically entails $y$, then giving up $x$ discards less

information than giving up $y$ (since retracting $y$ requires retracting $x$ as well, while $x$ can be retracted without retracting all of its logical consequences). To understand the motivation for *(EE3)*, notice that *(EE2)* requires that $x \wedge y \preceq_K x$ and $x \wedge y \preceq_K y$. Since the minimal retraction of $x \wedge y$ requires the retraction of either $x$ or $y$, $x \wedge y$ must be at least as entrenched as either $x$ or $y$. Since the entrenchment ordering applies to the entire language, *(EE4)* assigns all elements of the language not in $K$ minimal status in the ordering. *(EE5)* requires logical tautologies to be maximal elements in the ordering.

The epistemic entrenchment relation uniquely determines a contraction operation via the following definition:

$$y \in K_x^- \text{ iff } y \in K \text{ and either } x \preceq_K x \vee y \text{or} \models x.$$

Once again, the corresponding revision operator may be obtained via the Levi identity. Revision and contraction operators defined in this manner satisfy all of the relevant rationality postulates.

## 2.2 Belief bases

Nebel [39] discusses contraction operators on ***belief bases*** , which are finite sets of sentences instead of infinite deductively closed belief sets. The motivations for defining belief change on belief bases is twofold. First, operators defined on belief bases are computationally viable (they do not have to operate on infinite sets). Second, belief change operations on belief bases permit reason maintenance, while those on belief sets do not. The base contraction operator $\simeq$ defined as:

$$B \simeq x = \begin{cases} (\bigvee_{C \in (B \downarrow x)} C) \wedge (B \vee \neg x) & \text{if } \not\models x \\ B & \text{otherwise} \end{cases}$$

satisfies most, but not all, of the AGM postulates. Note that the term $(B \vee \neg x)$ ensures that the original belief base re-appears whenever $x$ becomes true. The corresponding revision operator can be defined, as before, via the Levi identity.

## 2.3 Other approaches

Frameworks for belief change are often distinguished on the basis of whether they adopt the *foundational* versus *coherentist* epistemology. The foundational approach requires that only facts having adequate justifications be accorded the status of beliefs. Thus, every belief must self-evident or have a non-circular, finite sequence of justifications grounded in a set of self-evident beliefs. The coherentist approach requires that minimal change be made to the original set of beliefs. The justification of an individual belief amongst a coherent set of beliefs is not is not its provability with respect to a set of self-evident axioms, but on the extent to which it coheres with all other beliefs. The AGM framework subscribes to the coherentist epistemology. Early work on belief change, some of it predating the AGM framework, focussed on the foundational approach, as exemplified in Doyle's TMS [15], de Kleer's ATMS [11] and the belief change system of Martins and Shapiro [35]. A number of approaches based on the foundational theory were developed concurrently with the AGM framework. These include the work of Dalal [9], Borgida [3], Winslett [52], Satoh [47], Weber [50], Fagin et al [16].

More recently, several studies have focussed on connections between nonmonotonic reasoning and belief change. Earliest among these is Rao and Foo's axiomatization of foundational and coherentist belief change using the language of auto-epistemic logic [45]. Brewka describes how belief change system can be built on a variant of the THEORIST system [6]. Similar results were later presented by Nebel [40] and Gärdenfors [20]. We have previously established connections between the process of contraction and the computation of extensions in PJ-default logic [23]. Boutilier establishes a similar connection in [4].

# Chapter 3

# A competence theory for belief change

## 3.1 Introduction

As with any problem with multiple competing solution techniques, belief change requires a good definition of competence, i.e., a specification of what it means to revise beliefs correctly. A competence theory must embody the consensus view of what constitutes ideal belief change. Like most computationally difficult problems, an idealization is not necessarily practical or implementable. Nonetheless, a formally well-specified best-case scenario serves as a yardstick, or a theoretical upper-limit, against which the competing solution techniques may be compared. Such a benchmark can also serve as a starting point for the development of new solution techniques, by making transparent the precise nature of the trade-offs made in effecting the transition from the idealization to the implementation.

In the state-of-the-art in belief change research, the definitions of rationality found in the AGM framework serve as a generally agreed upon specification of competence. We believe that these rationality postulates represent a useful first step, but are inadequate. The specific drawbacks are as follows:

- It is generally agreed that the *principle of informational economy* should guide any strategy for belief change. This requires that beliefs should be discarded as little as possible while effecting belief change. The AGM rationality postulates seek to satisfy this requirement, but with limited success. We shall show that, as a consequence of the belief representation scheme and an unduly narrow definition of what constitutes success for a belief change operation, beliefs may be unduly discarded by operators defined in the AGM framework.

- The AGM framework provides an inadequate account of the process of retracting a belief. Thus, the addition of a belief is duly recorded in the belief state of an agent, but the retraction of a belief is never recorded. We shall show that this can unduly restrict the space of candidate outcomes of a belief change operation.
- The AGM framework does not specify belief change beyond a single step.
- The AGM framework provides no prescription on how beliefs must change when the belief input is not fully credible. We believe that any approach to handling uncertain, or less credible, belief inputs should involve a generalization of techniques applied when the belief inputs are fully credible, instead of requiring a totally distinct set of techniques.

In attempting to address these drawbacks, we shall progressively generalize the AGM approach through the following four steps. The first three involve augmenting the belief representation scheme. The final step involves modifying the notion of success for a belief change operation.

In augmenting the belief representation scheme, we shall distinguish between the notions of the *belief state* and *commitment state* of an agent.

*Belief state:* The AGM framework views a belief state as denoting the set of beliefs currently held by an agent. We view a belief state as a representation of the beliefs that may potentially be held by an agent. Fundamentally, this shift is identical to the shift from classical logic to a nonmonotonic formalism for representing a knowledge base. This view admits the possibility that there might be several candidate beliefs that an agent is aware of, but only some that it actually commits to. A parallel observation in a nonmonotonic logic such as default logic [46] is the existence of multiple defaults in a knowledge base in the general case, only some of which become applicable at any given time.

*Commitment state:* Although a belief state may be viewed as a collection of all beliefs that an agent may potentially hold at any given time, an agent must reason with, or act upon, a single consistent set of beliefs. We shall refer to such a set as the commitment state of an agent. A given belief state may, in general, support several commitment states. Driven by the context, an agent must commit to a single element of the set of commitment states supported by its belief state, at any given point in time. We may draw a parallel between the commitment state of an agent and extensions in a nonmonotonic formalism such as default logic. Every extension is a valid commitment state supported by a belief state represented as a default theory, for a credulous reasoner. For a sceptical reasoner, a default theory supports only one commitment state, given by the intersection of all the extensions of the default theory.

Another crucial distinction is between the operations of revision and contraction. Recall that the AGM framework distinguishes between three possible epistemic attitudes to a belief $x$: $x$ is accepted, $x$ is undetermined and $x$ is rejected. While a revision operation causes a belief to be accepted or rejected, a contraction causes a belief to become undetermined. In the rest of this work, we shall treat revisions and contractions to be independent, symmetric operations at par with each other. It may be argued that contraction is only an intermediate step in a revision process via the Levi identity and has no status as an independent operation. To counter this claim, we shall point out that in most realistic scenarios, the beliefs of an agent represent an incomplete picture of the world (i.e. the agent is unable to commit to the truth or falsity of every sentence in the language). It makes sense, then, to talk of an operation that enforces the undetermined status of a belief. Clearly this cannot be achieved through a revision operation. An independent contraction operator must therefore be considered.

The steps we go through are as follows:

- A first step towards generalizing the belief representation scheme in the AGM framework is to identify a belief state with a set of theories as opposed to a single theory. Each theory represents a potential commitment state for an agent. We retain the AGM notion of success, so that every theory in the belief state obtained after a belief change operation satisfies the belief input (every theory contains the new belief after revision and every theory discards the retracted belief after contraction). This permits us to retain all beliefs that do not directly contradict the belief input. However, some beliefs are nevertheless irretrievably discarded. Contractions are not recorded in a belief state, and no prescription is given on how to handle uncertain or less credible beliefs inputs.

- The second generalization of the belief representation scheme introduces the notion of *disbeliefs*. Informally, a disbelief represents the dual of a belief. Thus, just as a belief may be viewed as a record of a revision transaction, a disbelief may be viewed as a record of a contraction transaction. The new scheme views a belief state as a set of theories representing beliefs that may potentially be held, together with one other theory representing the current set of disbeliefs. The disbelief theory contains the negations of the beliefs that may not be held. By requiring every theory in the set of theories representing the beliefs to be consistent with the disbelief theory, we can guarantee that the current set of disbeliefs do not appear in any consistent set of beliefs. Any element of the set of belief theories, together with the theory representing disbeliefs constitutes a candidate commitment state for an agent. The AGM notion of success is retained. As with the previous representation scheme, we are able to retain all beliefs that do not directly contradict the belief input. As before, some beliefs must nevertheless be discarded as a consequence of retaining the AGM notion of success. Contractions can be recorded in the theory representing the disbeliefs. However, when a contraction operation contradicts an earlier contraction (such as contract $a$ followed by contract $\neg a$) we are forced to discard the memory of one of these contractions, since we are committed to maintaining exactly one consistent theory denoting the current disbeliefs. As before, we are unable to account for uncertain or less credible belief inputs.

- The third generalization of the representation scheme involves viewing a belief state as a collection of theories denoting beliefs together with a collection of theories denoting disbeliefs. A pair consisting of a belief theory and a disbelief theory represents a candidate commitment state for the agent. Informally, when an agent is given a set of potential beliefs and a set of potential disbeliefs, it may take two complementary approaches to arriving at a maximal commitment state. In the first approach, it selects a maximal consistent subset of the set of beliefs, and subsequently identifies a maximal subset of the set of disbeliefs that is consistent with the belief set already selected. In the second approach, which is the dual of the first approach, it selects a maximal consistent subset of the set of disbeliefs and subsequently identifies a maximal subset of the set of beliefs which is consistent with the disbelief set already selected. This scheme permits us to minimize the discarding of disbeliefs, since contradictory contraction operations can be recorded in distinct disbelief theories. However, since the AGM notion of success is retained, beliefs as well as disbeliefs that directly contradict the current belief input must be discarded. No support is provided for uncertain or less credible belief inputs.

- The final step involves taking the representation scheme in the previous step and relaxing the AGM constraint for success. Under the redefined notion of success, a revision operation succeeds if there is at least one belief theory in the resulting belief state which contains the new belief. Similarly, a contraction operation succeeds if this is reflected in at least one disbelief theory in the resulting belief state. This provides us with a framework in which no belief or disbelief needs to be discarded. In effect, we obtain an idealization of the principle of informational economy. We set up a set of conditions that uniquely define belief change operations in this framework and argue that they provide a more appropriate definition of competence than the AGM postulates, specially for belief change systems based on nonmonotonic logics where one is not constrained to define mappings between single classical theories. We specify operators which satisfy these conditions. We establish formally our intuitive observation that the new framework strictly subsumes the AGM framework. Finally, we point out how uncertain or less credible belief inputs can be handled within the same integrated framework.

## 3.2 A critique of AGM rationality

The AGM framework defines a representation scheme for the belief state of an agent, a repertoire of epistemic inputs that drive belief change operations, a repertoire of belief change operations, a set of rationality requirements for these operations and a set of operators for each of these operations. We shall argue in this section that while the AGM framework represents an important step towards a uniform and principled treatment of the dynamics of belief states, it has several shortcomings as well. We will show that most of these shortcomings are a consequence of inadequacies in the AGM representation scheme for an agent's belief state, the AGM formulation of belief change operations and the AGM constructions of operators for these operations.

In the rest of this section we shall refer to AGM *knowledge sets* interchangeably as *belief states*, pending our new definition of a representation of a belief state.

Our critique of the AGM framework consists of five major arguments.

*(A) Non-compliance with the informational economy requirement.*

The guiding theme in the AGM rationality postulates is ensuring *minimal change* while mapping one belief state to another in response to an epistemic input. Yet AGM-rational belief change operators cause beliefs to be irretrievably discarded. Consider the following scenario. At belief state $K$, we discover that $x$ is true in the world and accordingly revise our beliefs to obtain the new belief state $K^*_x$. We are then told that the previous revision step was incorrect - that there was, in fact, insufficient evidence to conclude that $x$ was true in the world. We must therefore retract our belief in $x$ to obtain a new belief state $(K^*_x)^-_x$. Ideally, our misadventure into believing $x$ should cause no lasting damage, i.e., we should get back at least those beliefs that we started off with. Formally, we may state this requirement as follows:

$$S \subseteq (S^*_x)^-_x$$

where $S$ is some as yet unspecified representation of a belief state. In later sections, we shall explore a space of possible representations in order to eventually satisfy this requirement. This requirement is not satisfied in the AGM framework. The following example shows that there exist AGM-rational operators $*$ and $-$ which violate this requirement.

**Example 1** Consider a propositional language with the alphabet given by the singleton set $\{a\}$. Let $K = Cn(\neg a)$. There is exactly one way in which $K$ can be revised with the new belief $a$ using an AGM-rational revision operator $*$, yielding the outcome $K^*_a = Cn(a)$. Subsequent retraction of $a$ from this belief state yields $(K^*_a)^-_a = Cn(\top)$. Clearly, $K \not\subseteq (K^*_a)^-_a$. □

In this example, the problem with the operator $*$ irretrievably discarding the belief $\neg a$ stems from the AGM framework requiring that a belief state be represented as a single consistent theory and forcing the resulting belief state to reflect the changes required by the belief input. Any operator which has to accommodate a new belief that contradicts existing beliefs and maintain consistency at the same time, must necessarily discard some beliefs.

Consider another aspect of the problem of discarding beliefs. At belief state $K$, we revise with the belief $x$ and subsequently discover that we were wrong and that, in fact, $\neg x$ is true in the world. Once again, we may argue that there should be no lasting damage, i.e.

$$S \subseteq (S^*_x)^*_{\neg x}$$

where is some appropriate representation of a belief state. The following example shows that even this requirement is not satisfied in the AGM framework. The problem in this example stems from the inability of the AGM framework to retain all of the candidate outcomes of a belief change operation.

**Example 2** Let $\{a, b\}$ be the alphabet of our language. Let the initial belief state be:

$$K0 = Cn(\{a, b\})$$

First, we revise $K0$ with $\neg b$.

$$K0 \downarrow b = \{O1, O2\} \text{ where}$$
$$O1 = Cn(\{\neg a \vee b, \neg b \vee a\})$$
$$O2 = Cn(\{a\})$$

We shall use an AGM-rational partial meet contraction operator where a transitively relational selection function $S$ returns $O1$.

$$S(K0 \downarrow b) = O1$$

Then:

$$K0^-_b = O1$$
$$K1 = K0^*_{\neg b} = (K0^-_b)^+_{\neg b} = Cn(O1 \cup \{\neg b\}) = Cn(\{\neg a, \neg b\})$$

We now revise $K1$ with $b$.

$$K1 \downarrow \neg b = \{O1', O2'\} \text{ where}$$
$$O1' = Cn(\{\neg a\})$$
$$O2' = Cn(\{\neg b \vee a, \neg a \vee b\})$$

Once again, an AGM-rational partial meet contraction operator exists with a transitively relational selection function $S'$ which returns $O1'$.

$$S''(K1 \downarrow \neg b) = O1'$$

Then:

$$K1^-_{\neg b} = O1'$$
$$K2 = K1^*_b = (K1^-_{\neg b})^*_b = Cn(\{\neg a, b\})$$

Clearly:

$$K0 \not\subseteq (K0^*_{\neg b})^*_b$$

In the next section, we shall show that a more general belief representation scheme allows us to retain beliefs that would otherwise be discarded.

*(B) Lack of an explicit representation for contracted beliefs.*

The limited expressive power of the belief representation scheme in the AGM framework can result in crucial inputs to the belief change function being ignored. Consider the following example.

**Example 3** Let $\{b, f\}$ be the alphabet of our language. Let the initial belief state be:

$$K0 = Cn(\{b \rightarrow f\})$$

After contracting $f$ using an AGM-rational operator $-$, we get the following outcome:

$$K1 = K0^-_f = Cn(\{b \rightarrow f\})$$

Since $f$ is not a consequence of the beliefs in $K0$, no change is made to $K0$. Revising $K1$ with $b$ results in the belief state:

$$K2 = K1^*_b = Cn(\{b, b \rightarrow f\})$$

Thus, the agent starts believing $f$ again, although the only new information (the belief $b$) obtained since being told to retract the belief $f$ does not in itself require that $f$ be believed again. A more detailed analysis reveals that when $K1$ is revised with $b$, three different entities need to be considered:

A: $b \rightarrow f$ and its consequences are believed.

B: $f$ is retracted.

C: $b$ is believed.

Prioritizing these entities informally using a relation $>$, where $x > y$ denotes that $x$ has higher priority over $y$, a variety of outcomes are possible. Note that the $*$ operator we use here is an idealized revision operator, not to be confused with an AGM revision operator. We list some of the possibilities below.

- If $C > A > B$ then $K1^*_b = Cn(\{b, b \rightarrow f\})$.
- If $A > C > B$ then $K1^*_b = Cn(\{b, b \rightarrow f\})$.
- If $C > B > A$ then $K1^*_b = Cn(\{b\})$.
- If $B > C > A$ then $K1^*_b = Cn(\{b\})$.
- If $B > A > C$ then $K1^*_b = Cn(\{b \rightarrow f\})$.
- If $A > B > C$ then $K1^*_b = Cn(\{b \rightarrow f\})$.

We do not list all the possibilities here, but clearly the three distinct entities and their relative prioritization need to be considered in generating an outcome. For instance, let $b$ denote that Tweety is a bird and $f$ denote that Tweety flies. Then, starting with a belief in an instance of the rule "birds fly", after retracting the belief that Tweety flies and then being told that Tweety is a bird, it seems reasonable to remove the "birds fly" rule from the status of a first-class belief, given new information regarding Tweety's flying ability and the fact that Tweety is a bird, giving a final belief state $Cn(b)$. This corresponds to the case where $C > B > A$. □

In the AGM framework, the effects of a revision are explicitly recorded and retained until brought into question by subsequently acquired beliefs. Contractions, however, are not explicitly recorded. Thus, while the belief state obtained as a result of a contraction operation is constrained not to contain the contracted belief, subsequent belief states have no memory of this contraction operation. Clearly, an explicit representation of contracted beliefs is necessary in the belief representation scheme.

*(C) The problem of spurious beliefs*

The lack of explicit representations of contractions in the AGM framework manifests itself in other forms of pathological behaviour as well. One of these is the appearance of unwarranted beliefs in a belief state; we shall refer to this as the *problem of spurious beliefs*. Consider the following example.

**Example 4** We shall use an AGM-rational system. Let $\{a, b, c\}$ be the alphabet of our language. The initial belief state is given by:

$$K0 = Cn(\{a, b\})$$

First, we shall contract $b$ from $K0$.

$$K0 \downarrow b = \{Cn(\{a, \neg c \vee b\}), Cn(\{a, c \vee b\}), Cn(\{\neg a \vee b, \neg c \vee b\}), Cn(\{\neg a \vee b, c \vee b\})\}$$

Let the outcome of applying a transitively relational selection function $S$ be:

$$S(K0 \downarrow b) = Cn(\{a, \neg c \vee b\})$$

Then:

$$K1 = K0^{-}_{b} = Cn(a, \neg c \vee b)$$

If we now revise $K1$ with $c$ (which is actually a trivial case of expansion), the outcome is:

$$K2 = K1^{*}_{c} = Cn(\{a, b, c\})$$

This is clearly unintuitive. We start by believing in $a$ and $b$ and their consequences. Subsequently, we find reason to disbelieve $b$. Finally, we find that we have reason to believe $c$. In itself, this does not provide sufficient grounds to start believing in $b$. Yet we end up in a belief state containing the spurious belief $b$.□

It may be argued that this is a consequence of the requirement that a belief state be a deductively closed theory in the AGM framework (we accept the belief $\neg c \vee b$ in

$K0$ in the previous example since it is a consequence of $b$). Relaxing the requirement for logical omniscience leads to systems that rely on belief bases for representing a belief state (such as Nebel's system [40]), but this has undesirable consequences, such as making belief changes syntax-dependent. We will show later that an explicit representation for disbeliefs can solve this problem too, without giving up logical omniscience.

*(D) Absence of an account for iterated belief change.*

Recall that the AGM framework provides a toolkit of four approaches to defining a contraction operator, and hence, via the Levi identity, a revision operator. All but one of these involve the use of some form of ordering to select one or more of the candidate outcomes in order to produce a final result. The exception is full meet revision/contraction, but revision in this style has the undesirable outcome of discarding all previous beliefs and retaining only the consequences of the new belief. Both partial meet and maxichoice revision/contraction involve the use of a selection function. In both cases, the definition of the selection function is parameterized by the current belief state $K$. In other words, a selection function $S$ applicable in the current belief state $K$ will not be applicable in a revised belief state $K'$. The AGM belief change operators provide no prescription of what the new selection function $S'$ applicable in belief state $K'$ should be. Similarly, the epistemic entrenchment relation is defined relative to the current belief state $K$ and the AGM belief change operators do not tell us how to obtain a new epistemic entrenchment relation for a revised belief state. The AGM operators only provide a specification of a belief change operation over a single step. Clearly, they provide an inadequate account of *iterated belief change*. This shortcoming of the AGM framework is a well-acknowledged one [49] [27] [38]. Suggested solutions include functions that map an epistemic entrenchment relation and an epistemic input to a new epistemic entrenchment relation [37].

*(E) The problem of requiring success.*

Fundamental to the AGM framework is the requirement that every belief change operation succeed. Thus, the revision postulate *(2*)* requires that the new belief be included in the revised belief state. Similarly, the contraction postulate *(4-)* requires that the contracted belief state not include the retracted belief, as long as the retracted belief is not a logical truth. However, it is possible to conceive of situations where the failure of a belief change operation might be warranted. This would be the case, for instance, if the credibility of the epistemic input were lower than that of existing beliefs. If we are told by an observer that an apple was spotted flying away from a tree instead of falling towards the ground, it is unlikely that we would discard our belief in Newton's laws of gravity. Let us formulate this situation as a revision operation. Starting with a belief state in which Newton's laws are accepted, and presented with an epistemic input consisting of the apple's observed behaviour (which contradicts currently accepted beliefs regarding the laws of gravity), we should ideally obtain a belief state in which we retain our belief in the laws of gravity, given the apparently dubious nature of the epistemic input. The revision operation should therefore fail. While remaining sceptical of this epistemic input, we would nonetheless incorporate the new belief in some way in our belief state (without accepting it) so that if we were to find out later that a gale-force storm was in progress at that time, we would actually accept the observer's account of the apple flying away. Clearly there are cases of belief change where the epistemic input is not accepted in the resulting belief state, but which causes some re-adjustment in the belief state nonetheless. Given the AGM framework definition of revision and contraction as operations which necessarily succeed, one option would be to expand the repertoire of belief change operations to include cases in which epistemic inputs of low credibility are accommodated. The problem with this option is that a single additional revision and contraction operator would not suffice. Epistemic inputs at different levels of credibility would have to be handled in different ways. We would thus need as many new operators as there would be levels of credibility. Another option is to generalize the AGM definition of revision and contraction. This is the approach we shall take in the next section.

## 3.3 A modified framework

In the previous section, we observed how inadequacies in the AGM belief representation scheme, in the AGM formulation of revisions and contraction and in the AGM revision and contraction operators limited the applicability of the AGM framework in most practical settings. We shall address the first two of these issues in this section by presenting a new system for representing an agent's belief state, and by reformulating the notion of revision and contraction. We shall explore a space of three possible approaches, which we shall refer to as:

- The sets of theories approach.
- The constrained sets of theories approach.
- The sets of constrained theories approach.

### Sets of theories

Recall our earlier criticism of the AGM rationality postulates for not representing an idealization of minimal belief change. This problem partly stems from the AGM-rational operators selecting some of the candidate outcomes of a belief change operation and discarding others, as illustrated in Example 2. This suggests a generalization of the AGM representation of a belief state along the following lines. As before, we shall view belief change as a mapping from a belief state and an epistemic input to another belief state. However, we shall view a belief state as a collection of deductively closed theories instead of a single deductively closed theory. Thus a belief state $S$ will be defined as $S = \{K_1, K_2, \ldots\}$ where each $K_i$ is an AGM knowledge set. Each $K_i$ represents the commitment state of an agent in belief state $S$. We shall dub this the *sets of theories approach.* We shall also generalize the definitions of revision and contraction to reflect this shift in representation. Let $R$ and $C$ be revision and contraction operators under our new definition.

*Revision:* Revision of a belief state $S$ with a belief $x$ results in a belief state $S^R_x$ such that for every $K_i \in S^R_x$, $x \in K_i$.

*Contraction:* Contraction of a belief $x$ from a belief state $S$ results in a belief state $S^C_x$ such that for every $K_i \in S^C_x$, $x \notin K_i$.

We may define the operator $C$ in terms of the AGM *removal operator* $\downarrow$. When $\not\models x$:

$$S^C_x = \{K' \mid K' \in (K \downarrow x), K \in S\}$$

In case $\models x$, $S^C_x = S$.

The revision operator is defined, as in the AGM framework, via the Levi identity.

$$S^R_x = \{Cn(K \cup \{x\}) \mid K \in S^C_{\neg x}\}$$

Note that most accounts of how an agent acts requires that an agent commit to a single consistent set of beliefs. However, it is also well-recognized that real-life agents reason using multiple contexts. In other words, an agent may hold several, potentially mutually inconsistent, sets of beliefs. In the sets of theories approach, an agent retains multiple, potentially mutually inconsistent contexts. To be able to act, the agent must be able to select one such belief set, as its commitment state. This is a theory preference problem. We do not need to commit to any single theory preference strategy here since the process of belief change is independent of how theory preference is performed. We might view the AGM approach to belief change as involving two separate tasks:

- *Belief maintenance:* This involves generating candidate outcomes that achieve the required change and retain as many beliefs as possible. We may view the $\downarrow$ operator as performing this task.

- *Theory preference:* Selecting one of the candidate outcomes, or some combination of them, as the final outcome. The application of the selection function in the different ways mentioned earlier achieves this.

The sets of theories approach, as well the others we shall discuss subsequently, make these two tasks orthogonal. The operators $R$ and $C$ perform belief maintenance, and the outcome of belief change becomes independent of the theory preference step. This allows the theory preference strategy to be context-dependent, with no constraints being imposed by the belief maintenance process. Belief change is reduced to belief maintenance.

Following the AGM postulates for contraction and revision, we can establish some properties of the $C$ and $R$ operators. To do so, we need to establish some properties of the removal operator. Some of the lemmas that follow involving the removal operator have been stated and proved in [19] and elsewhere, but we shall go through the exercise of establishing these results for expository purposes.

**Lemma 1** *For every $K' \in (K \downarrow x)$, $K' \subseteq K$.*

**Proof:** Follows from the definition of $\downarrow$.

**Lemma 2** *If $x \in K$, then for every $K' \in (K \downarrow x)$, and for every $y$, either $x \vee y \in K'$ or $x \vee \neg y \in K'$, but not both.*

**Proof:** Since $x \in K$, clearly both $x \vee y \in K$ and $x \vee \neg y \in K$. Assume the converse, i.e., for every $K' \in (K \downarrow x)$, $x \vee y \notin K'$ and $x \vee \neg y \notin K'$. Then both $K' \cup \{x \vee y\} \models x$ and $K' \cup \{x \vee \neg y\} \models x$. In other words, $K' \cup \{y\} \models x$ and $K' \cup \{\neg y\} \models x$. Then $K' \models x$, which is a contradiction. If both $x \vee y \in K'$ and $x \vee \neg y \in K'$, then too $K' \models x$. □

**Lemma 3** *If $x \in K$, then for every $K' \in (K \downarrow x)$, $K \subseteq Cn(K' \cup \{x\})$.*

**Proof:** Assume the converse, i.e., there exists $K' \in (K \downarrow x)$ such that $K \not\subseteq Cn(K' \cup \{x\})$. Then there must exist some $y$ such that $y \in K - Cn(K' \cup \{x\})$. Since $y \in K$, $y \vee \neg x \in K$. Since there exists no $z$ where $z \not\models x$ such that $z \wedge (y \vee \neg x) \models x$, $y \vee \neg x \in K'$. Then $y \in Cn(K' \cup \{x\})$, which is a contradiction. □

**Lemma 4** *If $\models x \leftrightarrow y$, then $(K \downarrow x) = (K \downarrow y)$.*

**Proof:** Let $K' \in (K \downarrow x)$. By Lemma 1, $K' \subseteq K$. By definition, for any $z \in K - K'$, $K' \cup \{z\} \models x$. Then $K' \cup \{z\} \models y$. Hence $K' \in (K \downarrow y)$, by definition.□

**Lemma 5** *[19] If $x, y \in K$, $(K \downarrow x \wedge y) = (K \downarrow x) \cup (K \downarrow y)$.*

**Proof:** We shall refer to [19] for the proof.

**Lemma 6** *If $y \in K$ and there exists some $K' \in (K \downarrow x)$ such that $y \notin K'$, then $\neg y \vee x \in K'$.*

**Proof:** By definition, if $y \in K$ and $y \notin K'$, then $K' \cup \{y\} \models x$. This means that $K' \models \neg y \vee x$.□

**Lemma 7** *If $K' \subseteq K$ and $x \notin K'$, then there exists some $K'' \in (K \downarrow x)$ such that $K' \subseteq K''$.*

**Proof:** Follows from the definition of $\downarrow$.

**Lemma 8** *If $y \in K$ and for all $K' \in (K \downarrow x)$, $y \notin K'$, then $y \models x$.*

**Proof:** If $y \not\models x$, then, by definition, there must exist some $L$ such that $\{y\} \cup L \in (K \downarrow x)$.□

**Lemma 9** *If $K \not\models x$, then $(K \downarrow x \vee y) = (K \downarrow y)$.*

**Proof:** Let $K' \in (K \downarrow y)$. Since $x \notin K$ and since $K'$ is the largest subset of $K$ which does not contain $y$, by definition, $K' \in (K \downarrow x \vee y)$. The reverse direction can be similarly shown.□

**Theorem 1** *For the $C$ operator defined as above:*

1. *For every $K \in S_x^C$, there exists $K' \in S$ such that $K \subseteq K'$.*
2. *If for some $K \in S$, $x \notin K$, then $K \in S_x^C$.*
3. *If $\not\models x$, then for every $K \in S_x^C$, $x \notin K$.*

4. *If* $x \in K$, *for some* $K \in S$, *then there exists some* $K' \in S_x^C$ *such that* $K \subseteq Cn(K' \cup \{x\})$.

5. *If* $\models x \leftrightarrow y$, *then* $S_x^C = S_y^C$.

6. *For any* $K \in S_x^C$ *and* $K' \in S_y^C$, *there exists some* $K'' \in S_{x \wedge y}^C$ *such that* $K \cap K' \subseteq K''$.

7. *If, for some* $K \in S_{x \wedge y}^C$, $x \notin K$, *then there exists some* $K' \in S_x^C$ *such that* $K \subseteq K'$.

**Proof:**

1. If $\not\models x$, then, by definition, for every $K \in S_x^C$, there exists some $K' \in S$ such that $K \in (K \downarrow x)$. Then, by Lemma 1, $K \subseteq K'$. If $\models x$, $S_x^C = S$, hence the result trivially holds.

2. If $\not\models x$, then, by definition, for every $K'$ such that $K' \in (K \downarrow x)$ where $K \in S$, $K' \in S_x^C$. In case $x \notin K$, $K \in (K \downarrow x)$. Thus $K \in S_x^C$. If $\models x$, the $S = S_x^C$, hence the result trivially holds.

3. If $\not\models x$, then, by definition, for every $K \in S_x^C$, $K \in (K' \downarrow x)$ for some $K'$. Clearly $x \notin K$.

4. Let $K' \in (K \downarrow x)$. If $\not\models x$, $K' \in S_x^C$ by definition. By Lemma 3, $K \subseteq Cn(K' \cup \{x\})$. If $\models x$, then $S_x^C = S$, hence the result trivially holds.

5. If $\not\models x$, let $K' \in S_x^C$. Then there exists some $K \in S$ such that $K' \in (K \downarrow x)$. Then $K' \in (K \downarrow y)$ by Lemma 4. Thus $K' \in S_y^C$. The reverse direction can be similarly shown. In case $\models x$, $S_x^C = S_y^C = S$, hence the result rivially holds.

6. If $\models x$, then $S_{x \wedge y}^C = S_y^C$. Then $K' \in S_{x \wedge y}^C$, hence the result holds trivially. Similarly, if $\models y$, then $S_{x \wedge y}^C = S_x^C$. Then $K \in S_{x \wedge y}^C$, hence the result holds trivially. If $\not\models x$ and $\not\models y$, then let $K \in (K''' \downarrow x)$ where $K''' \in S$. Two cases are possible:

- $x \wedge y \notin K'''$. Then $(K''' \downarrow x \wedge y) = \{K'''\}$. Then $K''' \in S^C_{X \wedge y}$. Since $K \subseteq K'''$ by Lemma 1, the result holds.
- $x \wedge y \in K'''$. Then, by Lemma 5, $K \in (K''' \downarrow x \wedge y)$ and hence $K \in S^C_{x \wedge y}$. Thus if we set $K'' = K$, the result holds for any choice of $K' \in S^C_y$.

7. If $\models x$, then $S^C_{x \wedge y} = S^C_y$. If $K \in S^C_y$, then there exists some $K'' \in S$ such that $K \in (K'' \downarrow y)$. Since $S^C_x = S$, $K'' \in S^C_x$. By Lemma 1, $K \subseteq K''$. If $\models y$, then $S^C_{x \wedge y} = S^C_x$. Then the result is trivially proved by setting $K' = K$. Consider when $\not\models x$ and $\not\models y$. By definition, there exists some $K'' \in S$ such that $K \in (K'' \downarrow x \wedge y)$. By Lemma 1 $K \subseteq K''$. Since $x \notin K$, there exists some $K' \in (K'' \downarrow x)$ such that $K \subseteq K'$ by Lemma 7. By definition, $K' \in S^C_x$. Hence proved.□

**Theorem 2** *For the operator R defined as above:*

1. *For every $K \in S^R_x$, $x \in K$.*
2. *For every $K \in S^R_x$, there exists some $K' \in S$ such that $K \subseteq Cn(K' \cup \{x\})$.*
3. *If for some $K \in S$, $\neg x \notin K$, then $Cn(K \cup \{x\}) \in S^R_x$.*
4. *If $K \in S^R_x$, $K \models \perp$ iff $\models \neg x$.*
5. *If $\models x \leftrightarrow y$, then $S^R_x = S^R_y$.*
6. *For every $K \in S^R_{x \wedge y}$, there exists $K' \in S^R_x$ such that $K \subseteq Cn(K' \cup \{y\})$.*
7. *If there exists $K \in S^R_x$ such that $\neg y \notin K$, then there exists $K' \in S^R_{x \wedge y}$ such that $Cn(K \cup \{y\}) \subseteq K'$.*

**Proof:**

1. By definition, for every $K \in S^R_x$, $K = Cn(K' \cup \{x\})$ for some K'. Hence $x \in K$.
2. By definition, for every $K \in S^R_x$, $K = Cn(K'' \cup \{x\})$ for some $K'' \in (K' \downarrow \neg x)$ where $K' \in S$. By Lemma 1, $K'' \subseteq K'$, hence $Cn(K'' \cup \{x\}) \subseteq Cn(K' \cup \{x\})$, hence $K \subseteq Cn(K' \cup \{x\})$.

3. If for some $K \in S$, $\neg x \notin K$, then $(K \downarrow \neg x) = \{K\}$. Then $K \in S^C_{\neg x}$, by definition. Then $Cn(K \cup \{x\}) \in S^R_x$, by definition.

4. $\rightarrow$ If $K \in S^R_x$ and $K \models \perp$, then let $K = Cn(K' \cup \{x\})$ where $K' \in S^C_x$. By definition, every $K' \in S^C_x$ is satisfiable. Then for $K \models \perp$, it must be true that $\models \neg x$.

   $\leftarrow$ If $\models \neg x$, then for every $K \in S^R_x$, $x \in K$ by (1), hence $K \models \perp$.

5. Let $K \in S^R_x$. Then, by definition, there exists some $K' \in S$ such that $K = Cn(K'' \cup \{x\})$ where $K'' \in (K' \downarrow \neg x)$. By Lemma 1, $K'' \in (K' \downarrow \neg y)$. Then, by definition, $Cn(K'' \cup \{y\}) \in S^R_y$, hence $Cn(K'' \cup \{x\}) \in S^R_y$, hence $K \in S^R_y$. The reverse direction can be similarly shown.

6. Let $K \in S^R_x$. Then, by definition, there exists $K'' \in S$ such that $K = Cn(K''' \cup \{x \wedge y\})$ where $K''' \in (K'' \downarrow \neg x \vee \neg y)$. Clearly $K''' \not\models \neg x$. Hence there exists some $K'''' \in (K' \downarrow \neg x)$ such that $K''' \subseteq K''''$. Then, if $K' = Cn(K'''' \cup \{x\})$, $K' \in S^R_x$. Clearly, $Cn(K''' \cup \{x \wedge y\}) \subseteq Cn(Cn(K''' \cup \{x\}) \cup \{y\})$.

7. Let $K \in S^R_x$ and $\neg y \notin K$. The there must exist some $K'' \in S$ such that $K''' \in (K'' \downarrow \neg x)$ and $K = Cn(K''' \cup \{x\})$. Clearly, $\neg x \notin K'''$. Then, by Lemma 7 and given that $\neg y \in K'''$, there exists some $K'''' \in (K \downarrow \neg x \vee \neg y)$ such that $K''' \subseteq K''''$. Then $Cn(K'''' \cup \{x\} \cup \{y\}) \in S^R_{x \wedge y}$. Let $Cn(K'''' \cup \{x\} \cup \{y\}) = K'$. Then $Cn(Cn(K''' \cup \{x\}) \cup \{y\}) \subseteq K'$. Thus $Cn(K \cup \{y\}) \subseteq K'$.

This reformulation addresses the problem identified in Example 2 as shown below.

**Example 5** Let the initial belief state be:

$S0 = \{K_1\}$ where

$K_1 = Cn(\{a, b\})$

First, we revise $S0$ with $\neg b$.

$S1 = S0^R_{\neg b} = \{Cn(O1 \cup \{\neg b\}), Cn(O2 \cup \{\neg b\})\} = \{Cn(\{\neg a, \neg b\}), Cn(\{a, \neg b\}\} = \{K'_1, K'_2\}$

where $O1$ and $O2$ are as defined in Example 2.

Next, we revise $S1$ with $b$.

$$K_1 \downarrow \neg b = \{O1', O2'\}$$
$$K_2 \downarrow \neg b = \{Cn(\{a\}), Cn(\{\neg a \vee b, a \vee b\})\} = \{O3', O4'\}$$

where $O1'$ and $O2'$ are as defined in Example 2. Then:

$$S2 = S1^R_b = \{Cn(O1' \cup \{b\}), Cn(O2' \cup \{b\}), Cn(O3' \cup \{b\}), Cn(O4' \cup \{b\})\} = \{K''_1, K''_2, K''_3, K''_4\}$$

It turns out that $K''_1 = K''_4$ and $K''_2 = K''_3$. Therefore:

$$S2 = \{K''_1, K''_2\} = \{Cn(\{\neg a, b\}), Cn(\{a, b\})\}$$

It is easy to see that

$$S0 \subseteq (S0^R_{\neg b})^R_b$$

□

In general we can show the following result.

**Theorem 3** *If there exists $K \in S$ such that $x \in K$, then there exists $K''' \in (S^R_{\neg x})^R_x$ such that $K \subseteq K'''$.*

**Proof:** Let $K \in S$ such that $x \in K$. For the result to hold, there must exist some $K'$, $K''$ and $K'''$ such that $K' \in (K \downarrow x)$, $K'' \in (Cn(K' \cup \neg x) \downarrow \neg x)$ and $K''' = Cn(K'' \cup x)$ with $K \subseteq K'''$. Let $z$ be some sentence such that $Cn(z) = K$. There must exist some $K'''$ such that $K \subseteq K'''$. To prove this, assume the converse. Thus there exists no $K'''$ such that $K''' \models z$. Then there exists no $K'' \in (Cn(K' \cup \neg x) \downarrow \neg x)$ such that $\neg x \vee z \in K''$. Two cases are possible:

1. $\neg x \vee z \in Cn(K' \cup \{\neg x\})$. Then, by Lemma 8, this is possible only if $\neg x \vee z \models \neg x$. Then $z \models \neg x$. But we know $z \models x$.

2. $\neg x \vee z \notin Cn(K' \cup \{\neg x\})$ for any $K' \in (K \downarrow x)$. Since $K \models z$, $\neg x \vee z \in K$. $\neg x \vee z \not\models x$. Then by Lemma 8 there must exist some $K' \in (K \downarrow x)$ where $\neg x \vee z \in K'$, which contradicts our assumption.

Thus there must exist some $K'''$ such that $K \subseteq K'''$.□

We can clearly minimize the discarding of beliefs by retaining all outcomes using the sets of theories approach, but this is an inadequate solution. Trivially, the problem identified in Example 1 remains. There are unique outcomes at every step in this case, yet beliefs are discarded.

As well, the problems arising from the lack of an explicit representation of contractions remain, as shown below.

**Example 6** Let us reformulate Example 3 in the sets of theories approach. The initial belief state is given by:

$S0 = \{K_1\}$ where
$K_1 = Cn(\{b \rightarrow f\})$

Contracting $f$ from $S0$, we get:

$$S1 = S0^C_f = \{K_1\}$$

Revising $S1$ with $b$, we get:

$$S2 = S1^R_b = \{Cn(\{b, b \rightarrow f\})\}$$

Clearly, retaining all outcomes in a belief state does not translate, even implicitly, to a memory of past contractions.□

The problem of spurious beliefs similarly remains.

## Constrained sets of theories

Example 6 suggests an augmentation of the sets of theories approach to include an explicit representation of contractions. In this new approach, which we shall dub the *constrained sets of theories approach*, a belief state is a pair $(T, D)$. $T$ is a set of deductively closed theories and corresponds to a belief state in the sets of theories approach. $D$ is a single deductively closed theory and corresponds to the current set of *disbeliefs*. We require that for every $K \in T$, $K \cup D \not\models \perp$. Informally, a disbelief corresponds to a contracted belief, so that $D$ is the theory that the agent is currently constrained to disbelieve. Thus, if an agent is currently constrained to disbelieve $x$, then $\neg x \in D$. $K_i$ together with $D$ constitutes a candidate commitment state for each $K_i \in T$. Thus a commitment state denotes the set of beliefs an agent commits to accept, together with a set of beliefs the agent commits to not accept, at any given point in time.

Let $C'$ and $R'$ be contraction and revision operators, respectively, in the constrained sets of theories approach.

*Revision:* Revision of a belief state $S = (T, D)$ with a belief $x$ results in a belief state $S_x^{R'} = (T_x^{R'}, D_x^{R'})$ such that for every $K \in T_x^{R'}$, $x \in K$.

*Contraction:* Contraction of a belief $x$ from a belief state $S = (T, D)$ results in a belief state $S_x^{C'} = (T_x^{C'}, D_x^{C'})$ such that $\neg x \in D_x^{C'}$.

In the case of contraction, notice that since every $K \in T_x^{C'}$ is required to be consistent with $D_x^{C'}$, it follows that $x \notin K$ for every $K \in T_x^{C'}$.

We may define the operators $C'$ and $R'$ in terms of the AGM removal operator $\downarrow$ and a selection function $f$ similar to the selection function used with the AGM maxichoice contraction operator.

The contraction operator $C'$ is defined as follows. If $\not\models x$:

$$S_x^{C'} = (T_x^{C'}, D_x^{C'}) \text{ where}$$
$$D_x^{C'} = Cn(f(D \downarrow x) \cup \{\neg x\})$$
$$T_x^{C'} = \{K' \mid K' \in (K \downarrow \neg D_x^{C'}), K \in T\}$$

In case $\models x$, the contraction operation fails, i.e., $S_x^{C'} = S$. Here, $\neg D_x^{C'}$ stands for the negation of the conjunction of the elements of $D_x^{C'}$.

To understand the motivations for the definition of the operator $C'$ given above, we need to consider the semantics of iterated contractions. In general, a successful contraction operation takes an agent from a belief state where the belief being contracted may be possibly held to a belief state where the belief is necessarily not held. In the case of systems where an agent may accept multiple belief sets in a given belief state, such as in the sets of theories approach and in the current approach, a successful contraction operation ensures that none of the beliefs sets in the resulting belief state contain the contracted belief. Contractions may contradict each other. For instance, the contraction of $x$ contradicts a prior contraction of $\neg x$, since it impossible to enforce disbelief in both $x$ and $\neg x$ (this equivalent to contracting the tautology $x \vee \neg x$). Therefore, the first step in a contraction operation with operator $C'$ is to obtain a new disbelief theory which contains $\neg x$ (by enforcing consistency with this theory, we can ensure that no theory in the resulting belief state contains $x$), where $x$ is the belief being contracted. This is achieved by revising the theory $D$ with $\neg x$, using an operator similar to revision in the AGM framework. First, $x$ is retracted from $D$ using maxichoice contraction, then the outcome is expanded with $\neg x$. We could define a similar operator in which $x$ is contracted from $D$ using partial meet contraction, which would avoid the well-recognized problem of obtaining complete theories as a consequence of revision using maxichoice contraction. However, the purpose of defining the operators in this section is purely expository; we wish demonstrate deficiencies in the constrained sets of theories approach to belief representation and motivate the need for an augmented representation. The deficiencies we identify are independent of how the disbelief theory is revised and are a consequence of the requirement that there be a unique disbelief theory (as well as the success requirement, which we shall address later). Nevertheless, we need to define operators in order to demonstrate the dynamics of belief change using this belief representation scheme. We therefore use the simpler maxichoice contraction operation as the basis for revising the disbelief theory. After the disbelief theory has been revised, maximal subsets of the current belief sets that are consistent with the new disbelief theory are identified, and these constitute the new set of belief sets.

Another aspect of iterated contractions that requires special mention is the generation of new disbeliefs in the absence of explicit contraction operations that require their addition, as shown in the following example.

**Example 7** Let the alphabet of our language be $\{a, b\}$. Let the initial belief state be given by:

$S0 = (T0, D0)$ where:
$D0 = Cn(\{a \vee \neg b\})$
$T0 = \{Cn(\{b\})\}$

After contracting $a$ from $S0$, we obtain:

$S1 = (T1, D1) = (T_a^{C'}, D_a^{C'})$ where:
$D1 = Cn(\{a \vee \neg b\} \cup \{\neg a\})$
$T1 = \{Cn(\{\neg a \vee b\})\}$

Notice that the belief $b$ is lost as a consequence of contracting $a$, given a prior contraction of $\neg a \wedge b$ (which shows up as the disbelief theory $D0$). This might appear anomalous, since the contracted beliefs do not in themselves require that $b$ be removed. Notice, however, that contracting $\neg a \wedge b$ translates to a requirement that the beliefs $\neg a$ and $b$ not be held simultaneously. When $a$ is subsequently contracted, this eliminates the possibility of enforcing disbelief in $\neg a$ (once again, this would translate to requiring disbelief in the tautology $a \vee \neg a$). Hence, the earlier requirement that $\neg a$ and $b$ be not held simultaneously now translates to a requirement that $b$ be not held. Hence $b$ is not contained in any theory in $T1$.□

Following the AGM postulates, we can establish the following properties of the $C'$ operator.

**Theorem 4** *For any operator $C'$ defined as above:*

1. *For every $K \in T_x^{C'}$, there exists some $K' \in T$ such that $K \subseteq K'$.*

*2. If $\not\models x$, then for every $K \in T_x^{C'}$, $x \notin K$.*

*3. If $\models x \leftrightarrow y$, then $S_x^{C'} = S_x^{C'}$.*

*4. If $x \in K$, for some $K \in T$, then there exists some $K' \in T_x^{C'}$ such that $K \subseteq Cn(K' \cup \{x\})$.*

**Proof:**

1. If $\models x$, the $S_x^{C'} = S$, and the result holds trivially. If $\not\models x$, then, by definition, for every $K \in T_x^{C'}$, there exists some $K' \in T$ such that $K \in (K' \downarrow \neg D_x^{C'})$. By Lemma 1, $K \subseteq K'$.

2. If $\not\models x$, by definition, for every $K \in T_x^{C'}$, $K \not\models \neg D_x^{C'}$. Since $D_x^{C'} \models \neg x$, then $K \not\models x$.

3. If $\models x$, then $\models y$ and vice versa. In either case, $S_x^{C'} = S_y^{C'} = S$. If $\not\models x$ or $\not\models y$, by Lemma 4 $D \downarrow x = D \downarrow y$. Since $f$ is a function, $D_x^{C'} = D_y^{C'}$. Hence $T_x^{C'} = T_y^{C'}$.

4. If $\models x$, then $S_x^{C'} = S$, hence the result trivially holds. Otherwise, let $K' \in (K \downarrow \neg D_x^{C'})$, where $S = (T, D)$, $K \in T$ such that $x \in K$. By definition, $D_x^{C'} = Cn(f(D \downarrow x)) \cup \{\neg x \wedge y\}$, where $y = Cn(f(D \downarrow x) \cup \{\neg x\}) - (Cn(f(D \downarrow x)) \cup Cn(\{\neg x\}))$ (taking $y$ to stand for the conjunction of its elements). We can rewrite $D_x^{C'}$ as $Cn(f(D \downarrow x)) \wedge \neg x \wedge y$ (taking each set of the form $Cn(X)$ to stand for the conjunction of its elements). Then $\neg D_x^{C'} = \neg Cn(f(D \downarrow x)) \vee \neg(\neg x \wedge y)$. Since $f(D \downarrow x) \subseteq D$ and $K \not\models \neg D$ by definition, $K \not\models \neg f(D \downarrow x)$. Then, by Lemma 9, $(K \downarrow \neg D_x^{C'}) = (K \downarrow \neg(x \wedge y)) = (K \downarrow (x \vee \neg y))$. Since $x \in K$, $x \vee \neg y \in K$. Then by Lemma 3, for every $K' \in (K \downarrow (x \vee \neg y))$, $K \subseteq Cn(K' \cup \{x \vee \neg y\})$. $Cn(K' \cup \{x \vee \neg y\}) \subseteq Cn(K' \cup \{x\})$, hence proved.

□

Notice that the reformulated versions of postulates (3−), (7−) and (8−) are not satisfied. We shall provide counter-examples for each of these cases. The following example shows that it is in general not true that if for some $S = (T, D)$, $K \in T$ and $x \notin K$, then $K \in T_x^{C'}$, hence a property similar to postulate (3−) cannot be proved.

**Example 8** Let the alphabet of our language be $\{x, y\}$. Let the initial belief state be given by:

$S0 = (T0, D0)$ where
$D0 = Cn(\{x \vee y\})$
$T0 = \{Cn(\{\neg y\})\}$

After contracting $x$ from $S0$, we obtain:

$S1 = (T1, D1) = (T0_x^{C'}, D0_x^{C'})$ where
$D0_x^{C'} = Cn(\{\neg x, x \vee y\})$
$T0_x^{C'} = \{Cn(\{\neg y \vee x\}), Cn(\{\neg y \vee x\})\}$

Notice that although $Cn(\{\neg y\}) \not\models x$, $Cn(\{\neg y\}) \notin T_x^{C'}$.□

The behaviour illustrated above stems from new disbeliefs being introduced as consequence of a sequence of contractions

The example below establishes the following two observations:

- It is not true, in general, that for any $K \in T_x^{C'}$ and $K' \in T_y^{C'}$, there exists some $K'' \in T_{x \wedge y}^{C'}$ such that $K \cap K' \subseteq K''$ (hence a property similar to postulate (7−) cannot be proved).

- It is not true, in general, that if for some $K \in T_{x \wedge y}^{C'}$, $x \notin K$, then there exists some $K' \in T_x^{C'}$ such that $K \subseteq K'$ (hence a property similar to postulate (8−) cannot be proved).

**Example 9** Let the alphabet of our language be $\{a, b\}$. Let the initial belief state be given by:

$S0 = (T0, D0)$ where:
$D0 = Cn(\{a, b\})$
$T0 = \{Cn(\{a, b\})\}$

Let the outcome of contracting $a$ from $S0$ be given by:

$S1 = (T1, D1) = (T0_a^{C'}, D0_a^{C'})$ where:
$D1 = Cn(f(D0 \downarrow a) \cup \{\neg a\}) = Cn(\{b, \neg a \vee b\} \cup \{\neg a\}) = Cn(\{b, \neg a\})$
$T1 = \{Cn(\{\neg a \vee b, \neg b \vee a\}), Cn(\{b\})\}$

Let the outcome of contracting $b$ from $S0$ be given by:

$S2 = (T2, D2) = (T0_b^{C'}, D0_b^{C'})$ where:
$D2 = Cn(f(D0 \downarrow b) \cup \{\neg b\}) = Cn(\{a\} \cup \{\neg b\}) = Cn(\{a, \neg b\})$
$T2 = \{Cn(\{\neg a \vee b, \neg b \vee a\}), Cn(\{a\})\}$

Let the outcome of contracting $a \wedge b$ from $S0$ be given by:

$S3 = (T3, D3) = (T0_{a \wedge b}^{C'}, D0_{a \wedge b}^{C'})$ where:
$D3 = Cn(f(D0 \downarrow a \wedge b) \cup \{\neg a \vee \neg b\}) = Cn(\{\neg a \vee b, \neg b \vee a\} \cup \{\neg a \vee \neg b\}) =$
$Cn(\{\neg a, \neg b\})$
$T3 = \{Cn(\{\neg a \vee b, \neg b \vee a\})\}$

Notice that $Cn(\{b\}) \in T1$, $Cn(\{a\}) \in T2$, $(b \vee a) \in T1 \cap T2$, yet there is no $K \in T3$ such that $K \models b \vee a$. This demonstrates the first of the two assertions.

Consider a different outcome of the contraction of $a$ from $S0$, corresponding to a different selection function $f1$.

$S4 = (T4, D4) = (T0_b^{C'}, D0_b^{C'})$ where:
$D4 = Cn(f1(D0 \downarrow a) \cup \{\neg a\}) = Cn(\{\neg a \vee b, \neg b \vee a\} \cup \{\neg a\}) = Cn(\{\neg a, \neg b\})$
$T4 = \{Cn(\{\neg a \vee b, \neg b \vee a\})\}$

As well, consider a different outcome of the contraction of $a \wedge b$ from $S0$, corresponding to this distinct selection function $f1$.

$S5 = (T5, D5) = (T0_{a \wedge b}^{C'}, D0_{a \wedge b}^{C'})$ where:
$D5 = Cn(f1(D0 \downarrow a \wedge b) \cup \{\neg a \vee \neg b\}) = Cn(\{b\} \cup \{\neg a \vee \neg b\}) = Cn(\{b, \neg a\})$
$T5 = \{Cn(\{\neg a \vee b, \neg b \vee a\}), Cn(\{b\})\}$

Notice that $Cn(\{b\}) \in T5$, $a \notin Cn(\{b\})$, yet there is no $K \in T4$ such that $Cn(\{b\}) \subseteq K$.□

Operator $C'$ does not satisfy the reformulated versions of postulates (7−) and (8−) on account of the fact that the selection function, used in revising the disbelief theory to accommodate the negation of the belief being contracted, is not constrained to select maximal elements of the power set of $D$ independent of the belief being contracted (as is the case in maxichoice operators that are fully AGM-rational [19]). We speculate that if such a constraint were imposed, the reformulated versions of postulates (7−) and (8−) would be satisfied. We shall not undertake the exercise of formally establishing this since our intent in defining this contraction operator is purely expository. We shall show later that the constrained sets of theories approach has deficiencies that are independent of how revision and contraction operators are defined in this system.

The revision operator $R'$ is defined as follows:

$$S_x^{R'} = (T_x^{R'}, D_x^{R'}) \text{ where}$$
$$D_x^{R'} = Cn(f(D \downarrow \neg x) \cup \{x\})$$
$$T_x^{R'} = \{Cn(K') \mid K' \in \delta(((K \cup \{x\}) \downarrow \neg D_x^{R'}), x), K \in T\}$$

where

$$\delta(T, x) = \{K \mid K \in T, K \models x\}$$

Thus $\delta$ is function which takes a set of theories and a sentence and returns the subset of theories which contains the sentence.

The first step in a revision operation is to suitably modify the disbelief theory so that it contains the new belief and thus blocks belief in its negation. The second step takes each element of the set of belief theories, identifies maximal subsets of the theory union the new belief that are consistent with the new disbelief theory and contain the new belief.

**Theorem 5** *The operator $R'$ defined above satisfies the following properties:*

*1. For every $K \in T_x^{R'}$, $x \in K$.*

*2. For every $K \in T_x^{R'}$, there exists some $K' \in T$ such that $K \subseteq Cn(K' \cup \{x\})$.*

*3. If $K \in T_x^{R'}$, $K \models \perp$ if $\models \neg x$.*

*4. If $\models x \leftrightarrow y$, then $S_x^{R'} = S_y^{R'}$.*

**Proof:**

1. Follows from the definition of $R'$.

2. Let $K'' \in \delta(((K' \cup \{x\}) \downarrow \neg D_x^{R'}), x)$. Clearly, $K'' \subset (K' \cup \{x\})$. Let $K = Cn(K'')$. $K \subseteq Cn(K' \cup \{x\})$ and $K \in T_x^{R'}$. Hence proved.

3. $x \in K$ for every $K \in T_x^{R'}$. Hence proved.

The operator $R'$ does not satisfy the reformulated versions of postulates $(4*), (7*)$ and $(8*)$.

With an explicit representation of disbeliefs/contractions, we can avoid the problem identified in Example 4.

**Example 10** We shall reformulate Example 4 in the constrained sets of theories approach. The initial belief state is given by:

$S0 = (T0, D0)$ where
$T0 = \{K_1\}$ with $K_1 = Cn(\{a, b\})$
$D0 = \{\}$

First, we contract $b$ from $S0$.

$S1 = S0_b^{C'} = (T0_b^{C'}, D0_b^{C'}) = (T1, D1)$ where
$D0_b^{C'} = Cn(\{\neg b\})$
$T0_b^{C'} = \{Cn(\{a, \neg c \vee b\}), Cn(\{a, c \vee b\}), Cn(\{\neg a \vee b, \neg c \vee b\}), Cn(\{\neg a \vee b, c \vee b\})\}$

Next, we revise $S1$ with $c$.

$$S2 = S1_c^{R'} = (T1_c^{R'}, D1_c^{R'}) = (T2, D2) \text{ where}$$
$$D1_c^{R'} = Cn(\{\neg b \wedge c\})$$
$$T1_c^{R'} = \{Cn(\{a, c\}), Cn(\{\neg a \vee b, c\})\}$$

Notice that for no $K \in S2$ is it true that $K \models b$. The disbelief in $b$ persists after revision $c$ since belief in $c$ is not sufficient evidence to warrant renewed belief in $b$. □

The explicit representation of disbeliefs/contractions in the constrained sets of theories approach represents an improvement over sets of theories approach. However, some problems persist, as the following examples show.

**Example 11** We shall reformulate Example 3 in the constrained sets of theories framework. The initial belief state is given by:

$$S0 = (T0, D0) \text{ where}$$
$$D0 = Cn()$$
$$T0 = \{Cn(\{\neg b \vee f\})\}$$

Contracting $f$ from $S0$, we get:

$$S1 = S0_f^{C'} = (T0_f^{C'}, D0_f^{C'}) = (T1, D1) \text{ where}$$
$$D0_f^{C'} = Cn(\{\neg f\})$$
$$T0_f^{C'} = \{Cn(\{\neg b \vee f\})\}$$

Revising $S1$ with $b$, we get:

$$S2 = S1_b^{R'} = (T1_b^{R'}, D1_b^{R'}) = (T2, D2) \text{ where}$$
$$D1_b^{R'} = Cn(\{\neg f \wedge b\})$$
$$T1_b^{R'} = \{Cn(\{b\})\}$$

Notice that one of the candidate outcomes $Cn(\{b\}) \in T1_b^{R'}$ corresponds to the ordering $C > B > A$ or the ordering $B > C > A$ in Example 3. Once again, the other possibilities are not considered, since the contraction and revision operations succeed. We shall see later that an augmented belief representation system with operators which relax the success requirement permits us to consider all of the possible outcomes mentioned in Example 3. In this case though, the outcome $Cn(\{b, f\})$ does not appear since the disbelief in $f$ is explicitly recorded and retained. *Box*

The following example demonstrates another instance of how potential outcomes are eliminated from consideration on account of the success requirement.

**Example 12** Let $\{b\}$ be the alphabet of our language. Let the initial belief state be:

$$S0 = (T0, D0) \text{ where}$$
$$D0 = Cn()$$
$$T0 = \{Cn(\{b\})\}$$

After contracting $b$ from $S0$, we obtain:

$$S1 = S0_b^{C'} = (T0_b^{C'}, D0_b^{C'}) = (T1, D1) \text{ where}$$
$$D0_b^{C'} = Cn(\{\neg b\})$$
$$T0_b^{C'} = \{Cn()\}$$

Analyzing the problem along the lines of Example 3, we notice that when $b$ is contracted from $S0$, two different entities need to be considered.

- A: $b$ and its consequences are believed.
- B: $b$ is retracted.

Depending on how these entities are prioritized, two outcomes are possible:

- If $A > B$ then the outcome is $Cn(\{b\})$.
- If $B > A$ then the outcome is $Cn()$.

The constrained sets of theories approach generates only the first outcome, since the contraction operation is defined to always succeed. Consequently, beliefs may be irretrievably discarded, such as the belief $b$ in this example.

The primary drawback of the belief representation scheme in the constrained sets of theories approach is the requirement that there be a single disbelief theory. This results in some disbeliefs, i.e., the memory of some previous contractions, being irretrievably discarded.

**Example 13** Let $\{b\}$ be the alphabet of our language. Let the initial belief state be:

$$S0 = (T0, D0) \text{ where}$$
$$D0 = Cn()$$
$$T0 = \{Cn()\}$$

First, we retract $b$ from $S0$.

$$S1 = S0_b^{C'} = (T0_b^{C'}, D0_b^{C'}) = (T1, D1) \text{ where}$$
$$D0_b^{C'} = Cn(\{\neg b\})$$
$$T0_b^{C'} = \{Cn()\}$$

Next, we retract $\neg b$ from $S1$.

$$S2 = S1_{\neg b}^{C'} = (T1_{\neg b}^{C'}, D1_{\neg b}^{C'}) = (T2, D2) \text{ where}$$
$$D1_{\neg b}^{C'} = Cn(\{b\})$$
$$T1_{\neg b}^{C'} = \{Cn()\}$$

Notice that the memory of the contraction of $b$ is lost since we have no way of recording mutually inconsistent disbeliefs given our commitment in this framework to a single consistent theory as the representation of the current set of disbeliefs.□

## Sets of constrained theories

Example 13 suggests that we augment our representation of a belief state to permit multiple sets of disbeliefs in the same way that the earlier two approaches permit multiple sets of beliefs. We shall augment the constrained sets of theories approach accordingly to obtain a framework which we shall call the *sets of constrained theories approach.*

As before we view belief revision as a process that takes a belief state and a belief input and produces a new belief state as a result. We view a belief state as a collection of commitment states. Formally, a commitment state is a pair $(B, D)$ (where $B$ is a belief theory and $D$ is a disbelief theory) such that the following conditions should hold:

- $B = Cn(B)$.
- $D = Cn(D)$.
- $B \cup D \not\models \perp$.

Given a belief state $S$, we may refer to its *universe of beliefs*, $\sum_+(S)$, where:

$$\sum_+(S) = \{b \mid \text{there exists } (B, D) \in S \text{ s.t. } B \models b\}$$

The *universe of disbeliefs*, $\sum_-(S)$, of a belief state $S$ is defined as:

$$\sum_-(S) = \{b \mid \text{there exists } (B, D) \in S \text{ s.t. } D \models b\}$$

Thus $\sum_+(S)$ contains all beliefs that may be held in some commitment state in $S$. Similarly, $\sum_-(S)$ contains all disbeliefs that may be held in some commitment state in $S$.

Belief inputs may be of two kinds:

- *Revision:* Revising a belief state $S$ with a belief $x$ results in a belief state $S'$ such that there exists a belief set $(B, D) \in S'$ where $B \models x$.
- *Contraction:* Contracting a belief $x$ from a belief state $S$ results in a belief state $S'$ such that there exists a belief set $(B, D) \in S'$ where $D \models \neg x$.

Notice that neither revision nor contraction is guaranteed to succeed in the above definitions. Revision merely requires that the new belief state contain at least one belief set with a belief theory $B$ that contains the new belief. Similarly, contraction requires that the new belief state contain at least one belief set with a disbelief theory $D$ containing the negation of the contracted belief. We shall refer to these as the *weak success* requirements for revision and contraction. The intuition is that the belief sets in the new belief state which represent non-successful revisions (contractions) correspond to outcomes with belief inputs that are not fully credible. In general, a belief state may contain multiple belief sets representing non-successful outcomes. Each such belief set intuitively represents one way in which a belief input with less than full credibility is ranked relative to the other beliefs/disbeliefs. A more compelling reason for relaxing the success requirement is the ability to retain all beliefs and disbeliefs at all times, as we shall see later. We believe that a competence theory for belief change with informational economy as its guiding theme must necessarily achieve this property of *full preservation.*

In this section, we have made two significant changes to the constrained sets of theories approach. First, we have augmented the belief representation scheme to permit multiple disbelief theories. Second, we have relaxed the success requirement. We shall establish in the next two sections that this approach satisfies most of our requirements for a good definition of competence. It may be asked, however, if both changes were necessary and if only one wouldn't suffice. Clearly the shift in representation is required to permit us to maintain contradictory disbeliefs. The need to relax the success requirement is not so obvious. For ease of exposition, we shall address this question later, after operators in the current approach have been defined.

## 3.4 Full preservation

We may now formalize the notion of a *fully preserving belief change operation* to represent the idealized case of change operations that do not discard any beliefs or disbeliefs. This may then be used as the yardstick to measure the extent to which a belief change operator achieves minimal change.

**Fully preserving revisions:** Let a revision operation map a belief state $S$ to a new belief state $S'$, where $x$ is the new belief. If $x \models \perp$, then $S = S'$. Otherwise, the revision is *fully preserving* iff the following conditions hold:

*Revision-1:* $\sum_+(S) \subseteq \sum_+(S')$.

*Revision-2:* For every $z \in \sum_+(S') - \sum_+(S)$, there exists some $y \in \sum_+(S)$ s.t. $x \wedge y \models z$.

*Revision-3:* $\sum_-(S) = \sum_-(S')$.

*Revision-4:* $x \in \sum_+(S')$.

*Revision-5:* For every $(B_i, D_i) \in S'$, if there exists $B$ where $B_i \subset B \subseteq \sum_+(S')$, then at least one of the following must hold:

- $B$ is not satisfiable.
- $D_i$ is such that for all $D$ where $D_i \subset D \subseteq \sum_-(S')$, $D$ is not satisfiable and $B \wedge D_i \models \perp$

*Revision-6* For every $(B_i, D_i) \in S'$, if there exists $D$ where $D_i \subset D \subseteq \sum_-(S')$, then at least one of the following must hold:

- $D$ is not satisfiable.
- $B_i$ is such that for all $B$ where $B_i \subset B \subseteq \sum_+(S')$, $B$ is not satisfiable and $D \wedge B_i \models \perp$.

Notice that revision with an inconsistent belief fails, unlike the AGM framework, where the revision would produce an inconsistent belief state.

*Revision-1* requires that the universe of beliefs of the revised belief state contain the universe of beliefs of the previous belief state. In other words, no beliefs may be discarded as a result of revision. *Revision-2* requires that any new beliefs introduced as a result of revision must be restricted to the consequences of the new belief taken together with all prior beliefs. This guarantees that no spurious beliefs are introduced as a result of revision. *Revision-3* requires that the universe of disbeliefs remain unchanged as a result of revision. Thus, the revision operation is constrained to not introduce any disbeliefs. *Revision-4* requires that the new belief be a part of the new universe of beliefs. This guarantees *weak success*. *Revision-5* requires that for every belief set in the revised belief state, the belief theory $B_i$ must either be a maximal consistent subset of the new universe of beliefs or must be a maximal subset of the new universe of beliefs which is consistent with some choice of a maximal consistent subset of the new universe of disbeliefs. *Revision-6* requires that for every belief set in the revised belief state, the disbelief theory $D_i$ must be either a maximal consistent subset of the new universe of disbeliefs or a maximal subset of the new universe of disbeliefs that is consistent with some choice of a maximal consistent subset of the new universe of beliefs.

Conditions *Revision-1* through *Revision-4* specify the revised belief state. In effect, they require that the the revised belief state consist of a universe of beliefs which contains the previous universe of beliefs, the new belief, and all of their consequences, and a universe of disbeliefs which is identical to the universe of disbeliefs of the prior belief state. Together, *Revision-5* and *Revision-6* specify what constitutes a valid commitment state. Establishing an exact symmetry between beliefs and disbeliefs, an agent, given a universe of beliefs and a universe of disbeliefs, may choose to commit to a maximal consistent subset of the potentially inconsistent universe of beliefs, or a maximal subset of the potentially inconsistent universe of disbeliefs. Concomitant with the choice of a maximal consistent subset of the universe of beliefs is a maximal subset of the universe of disbeliefs that is consistent with this choice. Similarly, concomitant with the choice of a maximal consistent subset of the universe of disbeliefs is a maximal subset of the universe of beliefs that is consistent with this choice. Every distinct pair $(B, D)$ of this form constitutes a valid state of commitment for the agent.

Notice that every belief state obtained as a result of applying a revision operator which satisfies all of the full preservation conditions constitutes a valid belief state under the sets of constrained theories representation.

**Observation:** *If $S$ is a belief state obtained as a result of applying a revision operator which satisfies* Revision-1 *through* Revision-6, *then for every $(B, D) \in S$:*

- $B = Cn(B)$.
- $D = Cn(D)$.
- $B \wedge D \not\models \perp$.

Notice also that condition *Revision-4* causes the weak success requirement defined earlier to be satisfied as well.

**Observation:** *If $S$ is a belief state obtained as a result of revision with $x$, applying a revision operator which satisfies* Revision-1 *through* Revision-6, *then there exists some $(B, D) \in S$ such that $B \models x$.*

A similar set of conditions can be formulated for the contraction operation.

> **Fully preserving contractions:** Let a contraction operation map a belief state $S$ to a new belief state $S'$, where $x$ is the belief being contracted. If $\models x$, then $S = S'$. Otherwise, the contraction is *fully preserving* iff the following conditions hold:
>
> *Contraction-1:* $\Sigma_-(S) \subseteq \Sigma_-(S')$.
>
> *Contraction-2:* For every $z \in \Sigma_-(S') - \Sigma_-(S)$, there exists some $y \in \Sigma_-(S)$ s.t. $\neg x \wedge y \models z$.
>
> *Contraction-3:* $\Sigma_+(S) = \Sigma_+(S')$.
>
> *Contraction-4:* $\neg x \in \Sigma_-(S')$.
>
> *Contraction-5:* For every $(Cn(B_i), Cn(D_i)) \in S'$, there exists $B_i \subseteq \Sigma_+(S')$ such that if there exists $B$ where $B_i \subset B \subseteq \Sigma_+(S')$, then at least one of the following must hold:
>
> - $B$ is not satisfiable.
> - $D_i$ is such that for all $D$ where $D_i \subset D \subseteq \Sigma_-(S')$, $D$ is not satisfiable and $B \wedge D_i \models \perp$.
>
> *Contraction-6:* For every $(Cn(B_i), Cn(D_i)) \in S'$, there exists $D_i \subseteq \Sigma_-(S')$ such that if there exists $D$ where $D_i \subset D \subseteq \Sigma_-(S')$, then at least one of the following must hold:
>
> - $D$ is not satisfiable.
> - $B_i$ is such that for all $B$ where $B_i \subset B \subseteq \Sigma_+(S')$, $B$ is not satisfiable and $D \wedge B_i \models \perp$.

Notice that the contraction operation fails if the belief being contracted is a logical truth, as is the case with the AGM framework.

*Contraction-1* requires that the universe of disbeliefs of the contracted belief state include the universe of disbeliefs of the prior belief state. In other words, no disbeliefs may be discarded as a result of contraction. *Contraction-2* requires that any new disbeliefs introduced as a result of contraction be restricted to the consequences of the new disbelief taken together with all prior disbeliefs. This guarantees that no spurious disbeliefs are introduced as a result of contraction. *Contraction-3* requires that the universe of beliefs remain unchanged as a result of contraction. Thus, the contraction operation is constrained to not introduce any new beliefs. *Contraction-4* requires that the new disbelief be contained in the new universe of disbelief. This guarantees *weak success.* As with revision, *Contraction-1* through *Contraction-4* specify the contracted belief state. Together, they require that the contracted belief state consist of a universe of beliefs which is identical to the prior universe of beliefs and a universe of disbeliefs which consists of all prior disbeliefs, the new disbelief, together with all of their consequences. *Contraction-5* and *Contraction-6* are identical to *Revision-5* and *Revision-6* and serve to specify a valid state of commitment for an agent in the contracted belief state.

As with revision, a belief state obtained as a result of applying a contraction operator which satisfies all of the full preservation requirements constitutes a valid belief state under the sets of constrained theories representation.

**Observation:** *If a belief state $S$ is obtained as a result of applying a contraction operator which satisfies* Contraction-1 *through* Contraction-6*, then for every* $(B, D) \in S$*:*

- $B = Cn(B)$.
- $D = Cn(D)$.
- $B \wedge D \models \perp$.

Notice that the weak success requirement for contraction is satisfied as a consequence of condition *Contraction-4*.

**Observation:** *If a belief state $S$ is obtained as a result of contracting $x$, applying a contraction operator which satisfies* Contraction-1 *through* Contraction-6*, then there exists $(B, D) \in S$ such that $D \models \neg x$, and hence $B \not\models x$.*

The *principle of irrelevance of syntax* follows from the full preservation properties, since we work with deductively closed theories as opposed to belief bases, and since we take the new belief or disbelief to be a single sentence.

**Observation:** *If a revision operator $r$ satisfies Revision-1 through Revision-6, then $S_x^r = S_y^r$ if $\models x \leftrightarrow y$.*

**Observation:** *If a contraction operator $c$ satisfies Contraction-1 through Contraction-6, then $S_x^c = S_y^c$ if $\models x \leftrightarrow y$.*

## 3.5 Minimal change operators

It is possible to define operators which satisfy the requirement for full preservation. In this section we shall define two operators, the *minimal revision operator* and the *minimal contraction operator*, and formally establish their compliance to the full preservation property.

**Definition 1 (Minimal revision operator)** *Let $S_x^{MR}$ represent the minimal revision of a belief state $S$ with a belief $x$. If $x \models \bot$, then $S_x^{MR} = S$. Otherwise, $S_x^{MR}$ is defined as:*

$$S_x^{MR} = S_x^{MR-Belief} \cup S_x^{MR-Disbelief}$$

*where:*

$$S_x^{MR-Belief} = \{(Cn(B), Cn(D)) \mid B \in (\Sigma_+(S) \cup \{x\}) \downarrow \bot, D \subseteq \Sigma_-(S) \text{ s.t. } B \wedge D \not\models \bot \text{ and there exists no } D' \text{ s.t. } (D \subset D' \subseteq \Sigma_-(S)) \wedge (B \wedge D' \not\models \bot)\}$$

*and*

$$S_x^{MR-Disbelief} = \{(Cn(B), Cn(D)) \mid D \in \Sigma_-(S) \downarrow \bot, B \subseteq (\Sigma_+(S) \cup \{x\}) \text{ s.t. } B \wedge D \not\models \bot \text{ and there exists no } B' \text{ s.t. } (B \subset B' \subseteq (\Sigma_+(S) \cup \{x\})) \wedge (B \wedge D' \not\models \bot)\}$$

The operator $MR$ identifies maximal commitments states through two steps. First, it identifies *maximal belief commitment states* via the operator $MR-Belief$. $MR-Belief$ identifies commitment states where the belief theory is a maximal consistent subset of the prior universe of beliefs, union the new belief. The corresponding disbelief theory is a maximal subset of the prior universe of disbeliefs which is consistent with the belief theory already selected. Second, the $MR$ operator identifies *maximal disbelief commitment states* via the operator $MR-Disbelief$. $MR-Disbelief$ identifies commitment states where the disbelief theory is a maximal consistent subset of the prior universe of disbeliefs. The corresponding belief theory is a maximal subset of the prior universe of beliefs union the new belief which is consistent with the disbelief theory already selected.

**Definition 2 (Minimal contraction operator)** *Let $S_x^{MC}$ represent the minimal contraction of a belief $x$ from a belief state $S$. If $\models x$, then $S_x^{MC} = S$. Otherwise, $S_x^{MC}$ is defined as:*

$$S_x^{MC} = S_x^{MC-Belief} \cup S_x^{MC-Disbelief}$$

*where:*

$$S_x^{MC-Belief} = \{(Cn(B), Cn(D)) \mid B \in \Sigma_+(S) \downarrow\perp, D \subseteq (\Sigma_-(S) \cup \{\neg x) \text{ s.t. } B \wedge D \not\models\perp \text{ and there exists no } D' \text{ s.t. } (D \subset D' \subseteq (\Sigma_-(S) \cup \{\neg x\}) \wedge (B \wedge D' \not\models\perp)\}$$

*and*

$$S_x^{MC-Disbelief} = \{(Cn(B), Cn(D)) \mid D \in (\Sigma_-(S) \cup \{\neg x\}) \downarrow\perp, B \subseteq \Sigma_+(S) \text{ s.t. } B \wedge D \not\models\perp \text{ and there exists no } B' \text{ s.t. } (B \subset B' \subseteq \Sigma_+(S)) \wedge (B \wedge D' \not\models\perp)\}$$

The $MC$ operator is the dual of the $MR$ operator. First, it identifies maximal belief commitment states using the $MC - Belief$ operator. $MC - Belief$ identifies commitment states where the belief theory is a maximal consistent subset of the prior universe of beliefs. The corresponding disbelief theory is a maximal subset of the prior universe of disbeliefs union the new disbelief which is consistent with the belief theory already selected. Second, the $MC$ operator identifies maximal disbelief commitment states via the $MC - Disbelief$ operator. $MC - Disbelief$ identifies commitment states where the disbelief theory is a maximal consistent subset of the prior universe of disbeliefs union the new disbelief. The corresponding belief theory is a maximal subset of the prior universe of beliefs which is consistent with the disbelief theory already selected.

The following results establish that the $MR$ and $MC$ operators comply with the full preservation requirements. They also establish that $MR$ and $MC$ are canonical members of their respective classes. In other words, if any revision (contraction) operator satisfies the full preservation requirements, then it is equivalent to $MR$ ($MC$).

**Theorem 6** *The minimal revision operator satisfies the full preservation properties* Revision-1 *through* Revision-6.

**Proof:** Consider the revision of a belief state $S$ with $x$ to obtain a belief state $S'$.

*Revision-1:* Let us assume the converse, i.e., $\sum_+(S) \not\subseteq \sum_+(S')$. Then there exists some $y$ such that $y \in \sum_+(S)$ and $y \notin \sum_+(S')$. From the definition of $S_x^{MR-Belief}$, the only way $y \in \sum_+(S)$ and $y \notin B_i$ for some $(B_i, D_i) \in S_x^{MR-Belief}$ is if $y \models \perp$ by Lemma 8. But if $y \in \sum_+(S)$, then these must exist some $(B_i', D_i') \in S$ such that $y \in Cn(B_i')$. We know by definition that every $B_i'$ is satisfiable, hence this is a contradiction.

*Revision-2:* Let us assume the converse, i.e., there exists some $z \in \sum_+(S') - \sum_+(S)$ for which there exists no $y \in \sum_+(S)$ s.t. $x \wedge y \models z$. By definition of $S_x^{MR-Belief}$, every $z \in Cn(B)$ for some $B \subseteq \sum_+(S) \cup \{x\}$, hence this is a contradiction.

*Revision-3:* We shall consider two cases:

- Assume there exists some $y \in \sum_-(S)$ such that $y \notin \sum_-(S')$. From the definition of $S_x^{MR-Disbelief}$ and Lemma 8, we know this is possible only if $y \models \perp$. But for $y \in \sum_-(S)$, $y \in D_i$ for some $(B_i, D_i) \in S$. By definition, every $D_i$ is satisfiable, hence this is a contradiction.
- Assume there exists some $y \in \sum_-(S')$ such that $y \notin \sum_-(S)$. Then there exists some $(B_i, D_i) \in S'$ such that $y \in D_i$. Then there exists some $D_i' \subseteq \sum_-(S)$ such that $D_i = Cn(D_i')$. Clearly, $y \in Cn(D_i') - D_i$. This is possible only if there exists no $(B_i'', D_i'') \in S$ such that $Cn(D_i') \subseteq D_i''$. By definition, this is possible only if $D_i' \models \perp$, which is contradiction since we know that $D_i$ is satisfiable.

*Revision-4:* Since $x \not\models \perp$, there must exist $B \subseteq (\sum_+(S) \cup \{x\}) \downarrow \perp$ such that $x \in B$. Hence there exists $(Cn(B), Cn(D)) \in S'$ such that $x \in Cn(B)$. Hence $x \in \sum_+(S')$.

*Revision-5:* If $(B_i, D_i) \in S_x^{MR-Belief}$, then for every $B$ such that $B_i \subset B \subseteq \sum_+(S')$, $B$ is unsatisfiable. If $(B_i, D_i) \in S_x^{MR-Disbelief}$, then $D_i$ is such that for any $D$ where $D_i \subset D_i \subseteq \sum_-(S')$, $D$ is unsatisfiable and for every $B$ such that $B_i \subset B \subseteq \sum_+(S')$, $B \wedge D_i \models \perp$. Hence proved.

*Revision-6:* If $(B_i, D_i) \in S_x^{MR-Disbelief}$, then for every $D$ such that $D_i \subset D \subseteq \sum_-(S')$, $D$ is unsatisfiable. If $(B_i, D_i) \in S_x^{MR-Belief}$, then $B_i$ is such that for every $B$ such that $B_i \subset B_i \subseteq \sum_+(S')$, $B$ is unsatisfiable and for every $D$ such that $D_i \subset D \subseteq \sum_-(S')$, $B_i \wedge D \models \perp$. Hence proved.

**Theorem 7** *If any revision operator $F$ satisfies the full preservation properties* Revision-1 *through* Revision-6, *then for any belief state $S$ and any belief $x$, $S_x^F = S_x^{MR}$.*

**Proof:** Since both $F$ and $MR$ satisfy *Revision-3*, $\sum_-(S_x^F) = \sum_-(S_x^{MR})$. Since both $F$ and $MR$ satisfy *Revision-1* and *Revision-2*, $\sum_+(S_x^F) = \sum_+(S_x^{MR})$. In the rest of this proof we shall consistently use $\sum_-(S_x^{MR})$ in place of $\sum_-(S_x^F)$ and $\sum_+(S_x^{MR})$ in place of $\sum_+(S_x^F)$. We shall consider two cases:

- Assume that there exists some $(Cn(B), Cn(D)) \in S_x^F$ such that $(Cn(B), Cn(D)) \notin S_x^{MR}$. Since $F$ satisfies *Revision-4*, $B$ satisfies at least one of the following conditions:

  1. For any $B'$ such that $B \subset B' \subseteq \sum_+(S_x^{MR})$, $B'$ is unsatisfiable. Since $F$ satisfies *Revision-5*, $D$ must satisfy at least one of the following conditions:

     – If there exists some $D'$ such that $D \subset D' \subseteq \sum_-(S_x^{MR})$, $D'$ is unsatisfiable. This implies $B \wedge D' \models \perp$. Then, by definition, $(Cn(B), Cn(D)) \in S_x^{MR-Belief}$.

     – If there exists some $D'$ such that $D \subset D' \subseteq \sum_-(S_x^{MR})$, $B \wedge D' \models \perp$. Then, by definition, $(Cn(B), Cn(D)) \in S_x^{MR-Belief}$.

  2. $D$ is such that for any $D'$ where $D \subset D' \subseteq \sum_-(S')$, $D'$ is unsatisfiable and for any $B'$ such that $B \subset B' \subseteq \sum_+(S)$, $B' \wedge D \models \perp$. Then, by definition $(Cn(B), Cn(D)) \in S_x^{MR-Disbelief}$.

- Assume there exists some $(Cn(B), Cn(D)) \in S_x^{MR}$ such that $(Cn(B), Cn(D)) \notin S_x^F$. We can show that this not possible by following an identical line of reasoning as above, substituting $F$ and $MR$.

□

**Theorem 8** *The minimal contraction operator satisfies the full preservation properties* Contraction-1 *through* Contraction-6.

**Proof:** Given the exact duality of beliefs and disbeliefs, the proof follows the same steps as that for Theorem 6 with beliefs and disbeliefs appropriately reversed.□

**Theorem 9** *If any contraction operator $F$ satisfies the full preservation properties* Contraction-1 *through* Contraction-6*, then for any belief state $S$ and any belief $x$, $S_x^F = S_x^{MC}$.*

**Proof:** Once again, the proof follows the same steps as that for Theorem 7, with beliefs and disbeliefs appropriately reversed. □

**Example 14** Let us reformulate Example 1 in the new framework. The initial belief state is given by:

$$S0 = \{(Cn(\{\neg a\}), Cn())\}$$

After revising $S0$ with $a$, we obtain:

$$S1 = S0_a^{MR} = \{(Cn(\{\neg a\}), Cn()), (Cn(\{a\}), Cn())\}$$

Contracting $a$ from $S1$, we obtain:

$$S2 = S1_a^{MC} = \{(Cn(\{\neg a\}), Cn(\{\neg a\})), (Cn(\{a\}), Cn())\}$$

Notice that there exists a belief set $(B, D) \in (S0_a^{MR})_a^{MC}$ such that $Cn(\{\neg a\}) \subseteq B$.□

**Example 15** Let us reformulate Example 2. The initial belief state is given by:

$$S0 = \{(Cn(\{a, b\}), Cn())\}$$

After revising $S0$ with $\neg b$, we obtain:

$$S1 = S0_{\neg b}^{MR} = \{(Cn(\{a, b\}), Cn()), (Cn(\{a, \neg b\}), Cn()), (Cn(\{\neg a, \neg b\}), Cn())\}$$

Further revising $S1$ with $b$, we obtain:

$$S2 = S1_b^{MR} = \{(Cn(\{\neg a, b\}), Cn()), (Cn(\{a, b\}), Cn()), (Cn(\{a, \neg b\}), Cn()), (Cn(\{\neg a, \neg b\}$$

Notice that $S2$ contains a commitment state with the original belief theory $Cn(\{a, b\})$.

□

**Example 16** Let us reformulate Example 3. Let the initial belief state be given by:

$$S0 = \{(Cn(\{b \rightarrow f\}), Cn())\}$$

After contracting $f$ from $S0$, we obtain:

$$S1 = S0_f^{MC} = \{(Cn(\{b \rightarrow f\}), Cn(\{\neg f\}))\}$$

Revising $S1$ with $b$, we obtain:

$S2 = S1_b^{MR} = \{(Cn(\{b, b \rightarrow f\}), Cn(\neg f \vee b\})), (Cn(\{b, b \rightarrow f\}), Cn(\neg f \vee \neg b)),$
$(Cn(\{b\}), Cn(\{\neg f\})), (Cn(\{b \rightarrow f\}), Cn(\{\neg f\}))\}$

Notice that all of the outcomes (corresponding to the different ways in which the belief in $b$, belief in $b \rightarrow f$ and disbelief $f$ are ordered) are now considered by any theory preference procedure that the agent may employ.□

**Example 17** Let us reformulate Example 4. Let the initial belief state be given by:

$S0 = \{(Cn(\{a, b\}), Cn())\}$

After contracting $b$ from $S0$, we obtain:

$S1 = S0_b^{MC} = \{(Cn(\{a, b\}), Cn(\{\neg b \vee c, \neg b \vee a\})), (Cn(\{a, b\}), Cn(\{\neg b \vee c, \neg b \vee \neg a\})),$
$(Cn(\{a, b\}), Cn(\{\neg b \vee \neg c, \neg b \vee a\})), (Cn(\{a, b\}), Cn(\{\neg b \vee \neg c, \neg b \vee \neg a\})),$
$(Cn(\{a, b \vee c\}), Cn(\{\neg b\})), (Cn(\{a, b \vee \neg c\}), Cn(\{\neg b\}))\}$

Revising $S1$ with $c$, we obtain:

$S2 = S1_c^{MR} = \{(Cn(\{a, b, c\}), Cn(\{\neg b \vee c, \neg b \vee a\})), (Cn(\{a, b, c\}), Cn(\{\neg b \vee c, \neg b \vee \neg a\})),$
$(Cn(\{a, b, c\}), Cn(\{\neg b \vee \neg c, \neg b \vee a\})), (Cn(\{a, b, c\}), Cn(\{\neg b \vee \neg c, \neg b \vee \neg a\})),$
$(Cn(\{a, c\}), Cn(\{\neg b\}))\}$

If disbelief in $b$ is to be maintained after revision with $c$, then $(Cn(\{a, c\}), Cn(\{\neg b\})) \in S2$ would be the appropriate outcome. □

**Example 18** Let us reformulate Example 12. Let the initial belief state be given by:

$S0 = \{(Cn(\{b\}), Cn())\}$

After contracting $b$ from $S0$, we obtain:

$$S1 = S0_b^{MC} = \{(Cn(\{b\}), Cn()), (Cn(), Cn(\{\neg b\}))\}$$

Notice that the commitment state $(Cn(\{b\}), Cn()) \in S1$ corresponds to the preferred outcome if $A > B$. Similarly, the commitment state $(Cn(), Cn(\{\neg b\})) \in S1$ would be the preferred outcome if $B > A$. □

**Example 19** Let us reformulate Example 13. Let the initial belief state be given by:

$$S0 = \{(Cn(), Cn())\}$$

After contracting $b$ from $S0$, we obtain:

$$S1 = S0_b^{MC} = \{(Cn(), Cn(\{\neg b\}))\}$$

Further contracting $\neg b$ from $S1$, we obtain:

$$S2 = S1_{\neg b}^{MC} = \{(Cn(), Cn(\{\neg b\})), (Cn(), Cn(\{b\}))\}$$

Notice that of the two commitment states in $S2$, one maintains disbelief in $b$ while the other maintains disbelief in $\neg b$. We would select one or the other depending on how the two contraction operations are prioritized. □

## 3.6 Should success be weakened ?

In this section, we shall address step three of our four-step argument in motivating the sets of constrained theories framework. We shall consider the situation where the AGM notion of success is retained with the sets of constrained theories belief representation scheme. We shall demonstrate that several of the problems we have identified would persist in this approach, and that, in addition ot the augmentation of the belief representation scheme, our weakening of the success requirement is essential.

We shall first define new versions of the $MR$ and $MC$ operators which satisfy the AGM success requirement for revision and contraction. This can be done in a straightforward manner by "filtering out", from the belief state returned by the $MR$ operator, those commitment states in which the belief theory does not contain the new belief in the case of revision. In the case of contraction, we "filter out" from the belief state returned by the $MC$ operator those commitment states where the disbelief theory does not contain the new disbelief. We define the following two functions to perform this "filtering":

$$\delta_r(S, x) = \{(B, D) \mid (B, D) \in S, x \in B\}$$
$$\delta_c(S, x) = \{(B, D) \mid (B, D) \in S, \neg x \in D\}$$

We can now define the counterparts of $MR$ and $MC$, $MR'$ and $MC'$, which satisfy the AGM-style success requirement as follows:

$$S_x^{MR'} = \delta_r(S_x^{MR}, x)$$
$$S_x^{MC'} = \delta_c(S_x^{MC}, x)$$

We shall now reformulate each of the examples of the previous section using $MR'$ and $MC'$ in place of $MR$ and $MC$, respectively.

**Example 20** Let us reformulate Example 1 in the new framework. The initial belief state is given by:

$$S0 = \{(Cn(\{\neg a\}), Cn())\}$$

After revising $S0$ with $a$, we obtain:

$$S1 = S0_a^{MR'} = \{(Cn(\{a\}), Cn())\}$$

Contracting $a$ from $S1$, we obtain:

$$S2 = S1_a^{MC'} = \{(Cn(), Cn(\{\neg a\}))\}$$

Notice that there exists no belief set $(B, D) \in (S0_a^{MR'})_a^{MC'}$ such that $Cn(\{\neg a\}) \subseteq B$. Both the beliefs $a$ and $\neg a$ are irretrievably lost.□

**Example 21** Let us reformulate Example 2. The initial belief state is given by:

$$S0 = \{(Cn(\{a, b\}), Cn())\}$$

After revising $S0$ with $\neg b$, we obtain:

$$S1 = S0^{MR'}_{\neg b} = \{(Cn(\{a, \neg b\}), Cn()), (Cn(\{\neg a, \neg b\}), Cn())\}$$

Further revising $S1$ with $b$, we obtain:

$$S2 = S1^{MR'}_{b} = \{(Cn(\{\neg a, b\}), Cn()), (Cn(\{a, b\}), Cn())\}$$

Notice that $S2$ contains a commitment state with the original belief theory $Cn(\{a, b\})$. In this case, weakening the success requirement does not provide any additional advantage. □

**Example 22** Let us reformulate Example 3. Let the initial belief state be given by:

$$S0 = \{(Cn(\{b \rightarrow f\}), Cn())\}$$

After contracting $f$ from $S0$, we obtain:

$$S1 = S0^{MC'}_{f} = \{(Cn(\{b \rightarrow f\}), Cn(\{\neg f\}))\}$$

Revising $S1$ with $b$, we obtain:

$$S2 = S1^{MR'}_{b} = \{(Cn(\{b, b \rightarrow f\}), Cn(\neg f \vee b\})), (Cn(\{b, b \rightarrow f\}), Cn(\neg f \vee \neg b)),$$
$$(Cn(\{b\}), Cn(\{\neg f\}))\}$$

Recall that we would like to be offered a choice of three candidate outcomes: $Cn(\{b\})$, $Cn(\{b, b \rightarrow f\})$ and $Cn(\{b \rightarrow f\})$. In this case, the last outcome, namely $Cn(\{b \rightarrow f\})$, is eliminated, on account of enforcing the success of the final revision operation. □

**Example 23** Let us reformulate Example 4. Let the initial belief state be given by:

$$S0 = \{(Cn(\{a, b\}), Cn())\}$$

After contracting $b$ from $S0$, we obtain:

$$S1 = S0^{MC'}_{b} = \{(Cn(\{a, b \vee c\}), Cn(\{\neg b\})), (Cn(\{a, b \vee \neg c\}), Cn(\{\neg b\}))\}$$

Revising $S1$ with $c$, we obtain:

$S2 = S1_c^{MR'} = \{(Cn(\{a,b,c\}), Cn(\{\neg b \vee c, \neg b \vee a\})), (Cn(\{a,b,c\}), Cn(\{\neg b \vee c, \neg b \vee \neg a\})),$
$(Cn(\{a,b,c\}), Cn(\{\neg b \vee \neg c, \neg b \vee a\})), (Cn(\{a,b,c\}), Cn(\{\neg b \vee \neg c, \neg b \vee \neg a\})),$
$(Cn(\{a,c\}), Cn(\{\neg b\}))\}$

Notice that we have fewer commitment states to chose from when we apply the $MC'$ operator, than when we apply the $MC$ operator. However, the final outcome is the same in both cases since revision with $c$ succeeds in every commitment state (there are no beliefs in the universe of beliefs which contradict $c$). □

**Example 24** Let us reformulate Example 12. Let the initial belief state be given by:

$S0 = \{(Cn(\{b\}), Cn())\}$

After contracting $b$ from $S0$, we obtain:

$S1 = S0_b^{MC'} = \{(Cn(), Cn(\{\neg b\}))\}$

Notice that the commitment state $(Cn(\{b\}), Cn()) \notin S1$ unlike the case with the $MC$ operator. Enforcing the AGM success requirement translates into a recency-based heuristic for ranking belief inputs. Thus, necessarily, the only outcome we get is the one corresponding to $B > A$. □

**Example 25** Let us reformulate Example 13. Let the initial belief state be given by:

$S0 = \{(Cn(), Cn())\}$

After contracting $b$ from $S0$, we obtain:

$S1 = S0_b^{MC'} = \{(Cn(), Cn(\{\neg b\}))\}$

Further contracting $\neg b$ from $S1$, we obtain:

$S2 = S1_{\neg b}^{MC'} = \{(Cn(), Cn(\{b\}))\}$

Any memory of the contraction of $b$ is erased as a result of enforcing the success of the contraction of $\neg b$. □

The previous examples demonstrate that the sets of constrained theories representation scheme alone is not sufficient to obtain an idealized system which satisfies the principle of informational economy. It is also necessary to weaken the AGM success requirement, as has been done in the case of the $MR$ and $MC$ operators. Another compelling reason for wanting to weaken the AGM success requirement is to provide support for uncertain or less-credible belief inputs. We shall address this question in a later section.

## 3.7 Connections with the AGM framework

In this section, we shall show that any system based on operators which satisfy the full preservation condition is a generalization of an AGM-rational system, in the sense that, starting with the same belief state, the set of outcomes generated by an operator such as $MR$ or $MC$ is a superset of the set of outcomes generated by an AGM-rational revision or contraction operator. Since the success requirement is weakened, fully preserving revision or contraction operators do not satisfy even reformulated versions of the AGM postulates. However, the informational economy requirement, which is the guiding theme of the AGM postulates, is satisfied in a stronger sense, as has been demonstrated in the previous sections. As well, we have seen that the principle of irrelevance of syntax is satisfied.

We shall require the following two results from [19] for proving our results. We shall restate them below in slightly simplified terms.

**Theorem 10 ([19])** *If a contraction operator satisfies postulates* $(1-)$ *through* $(8-)$ *and the expansion operator is defined as* $K^+_x = Cn(K \cup \{x\})$*, then the corresponding revision operator defined via the Levi identity satisfies postulates* $(1*)$ *through* $(8*)$.

Recall that a maxichoice contraction operator is defined such that $K^-_x = S(K \downarrow x)$ in case $\not\models x$ ($K^-_x = K$ otherwise). An *orderly* maxichoice operator is one which selects a $K_x^-$ such that there is a partial order $\leq$ on the power set of $K$ such that $K' \leq K_x^-$ for all $x$ and all $K' \in (K \downarrow x)$.

**Theorem 11 ([19])** *An orderly maxichoice contraction operator satisfies* (1∗) *through* (8∗).

**Theorem 12** *For a belief state $S = \{(B, D)\}$, and for any belief $x$ such that $x \not\models \perp$, there exists some $(B', D') \in S_x^{F1}$ such that $B' = B_x^{F2}$, where $F1$ is an operator that satisfies the full preservation requirements for revision and $F2$ is an operator that satisfies the AGM postulates for revision.*

**Proof:** We know from Theorem 6 that $F1$ is equivalent to the operator $MR$, hence we can consider $MR$ interchangeably with $F1$. Consider $S_x^{MR-Belief}$. Here, $\sum_+(S) = B$. Taking the converse of Lemma 8, since $x \not\models \perp$, there exists some $B'' \in (B \cup \{x\}) \downarrow \perp$ such that $x \in B''$. By definition, $B'' - \{x\} \in (B \downarrow \neg x)$. Consider a partial ordering $\leq$ defined on the power set of $B$ such that $B''' \leq B'' - \{x\}$, for all $x$ and for all $B''' \in (B \downarrow \neg x)$. Let $B_{\neg x}^- = B'' - \{x\}$. Then, by Theorem 11, the contraction operator satisfies postulates (1−) through (8−). Then $B_x^* = (B_{\neg x}^-)_x^+ = Cn((B'' - \{x\}) \cup \{x\}) = Cn(B'')$, where $*$ is a revision operator which satisfies revision postulates (1∗) through (8∗). We know from the definition of $S_x^{MR-Belief}$ that $(B', D') \in S_x^{MR}$, where $B' = Cn(B'')$, given an appropriate choice of $D'$.□

**Theorem 13** *For a belief state $S = \{(B, D)\}$, and for any belief $x$, there exists some $(B', D') \in S_x^{F1}$ such that $B' = B_x^{F2}$, where $F1$ is an operator that satisfies the full preservation requirements for contraction and $F2$ is an operator that satisfies the AGM postulates for contraction.*

**Proof:** We know from Theorem 8 that $F1$ is equivalent to the operator $MC$, hence we can consider $MC$ interchangeably with $F1$. If $\models x$, then $S_x^{MC} = S$. By definition, for an orderly maxichoice contraction operator $-$, $B_x^- = B$. Hence proved, by Theorem 11.

Consider the case when $\not\models x$. In computing $S_x^{MC-Disbelief}$, $\sum_-(S) = D$. Taking the converse of Lemma 8, since $\neg x \not\models \perp$, there exists some $D'' \in (D \cup \{\neg\}) \downarrow \perp$ such that $\neg x \in D''$. Consider some $b'' \in (B \downarrow x)$. Since $B'' \subseteq B$, $B \wedge D \not\models \perp$, $(D'' - \{\neg x\}) \subseteq D$ and $B'' \not\models x$, $B'' \wedge D'' \not\models \perp$. Since $\sum_+(S) = B$, $B'' \subseteq \sum_+(S)$. By definition, there exists no $B'''$ such that $B'' \subset B''' \subseteq \sum_+(S)$ and $B''' \wedge D'' \not\models \perp$. Let $B' = Cn(B'')$, $D' = Cn(D'')$. Clearly, $(B', D') \in S_x^{MC}$. Consider a partial ordering $\leq$ defined on the power set of $B$ such that $B''' \leq B''$, for all $x$ and for all $B''' \in (B \downarrow x)$. Let $B_x^- = B''$. Then by Theorem 11, the contraction operator $-$ satisfies contraction postulates $(1-)$ through $(8-)$. □

## 3.8 Discussion

This work has been motivated by several lacunae in the definition of competence for belief change provided by the AGM framework.

First, the rationality postulates for belief change are motivated by the requirement for informational economy, or minimal change, yet operators which satisfy these postulates discard beliefs, often irretrievably. This behaviour is caused, first, by a focus on mapping a single theory to a single other theory, and second, by the requirement that a belief change operation must necessarily succeed. We show that by generalizing the notion of belief change to a mapping between sets of theories, and by introducing a weaker notion of success, we can achieve a system where no beliefs are ever discarded. Since informational economy is the guiding theme in any definition of competence for belief change, and since this definition of competence must specify the ideal case, our framework qualifies as a competence theory for belief change in the general case, while the AGM framework does not. The AGM framework provides an adequate definition of competence only in the specific case when belief change is viewed as a mapping between single theories. However, a single classical theory is rarely an adequate representation for an agent's belief state, in the same way that classical logic is rarely an adequate representation language for a knowledge base.

In the case of a knowledge base, incomplete knowledge about the world, as well the defeasible nature of the inferences required, requires that some form of nonmonotonic logic be used as a representation language. Similarly, an agent typically needs to maintain multiple, mutually inconsistent sets of beliefs and selects one of these as the current commitment state depending on the reasoning context. We have provided a definition of competence when belief states have this richer structure.

Second, the AGM framework provides an inadequate account of contractions beyond a single step. Thus, while a revision is explicitly recorded in a belief state, an agent effectively loses all memory of a contraction from the very next step. We have shown how this can remove some candidate outcomes from consideration. Motivated by this observation, we have introduced an explicit representation of contractions, or disbeliefs in a belief state. In effect, we have introduced an exact symmetry between beliefs and disbeliefs. Thus, while the informational economy requirement seeks to minimize discarding memories of past revisions (or beliefs), we have interpreted it more generally to minimize discarding of memories of past contractions (or disbeliefs). The informational economy principle is motivated by the observation that it is expensive to acquire and maintain beliefs [19]. Arguably, it is as expensive to acquire and maintain evidence for an assertion (a belief), as it is to acquire and maintain evidence against an assertion (a disbelief). Therefore, we maintain multiple, mutually inconsistent sets of disbeliefs, in the same way that we maintain multiple, mutually inconsistent sets of beliefs in a belief state.

Third, the AGM framework provides no prescription for iterated belief change. Intrinsic to operators in the AGM framework is a choice function which selects among multiple candidate outcomes of a belief change operation. The problem stems from the lack of a specification of what the new choice function should be after a belief change operation. We solve this problem by obviating the need for choice. Every candidate outcome is retained in our framework, hence belief change in the iterated case is well-defined.

Fourth, the AGM framework provides no specification of how uncertain, or less credible, belief inputs are to be handled. By weakening the AGM notion of success, we are able to handle uncertain belief inputs within the same general framework, without having to define a new set of operators for this purpose. We shall discuss this in more detail later in this section.

In our model, a belief state is viewed as a collection of commitment states, where each commitment state is a pair consisting of a theory denoting the current beliefs and a theory denoting the current disbeliefs. Revision is simply the addition of a new belief to a belief state. Nothing is discarded. Thus, if belief in $\neg x$, or disbelief in $x$, is part of some commitment state in the belief state, neither is discarded as a result of revision with $x$. Contraction, similarly, involves the addition of a disbelief to a belief state. Notice that this is distinct from removing a belief, which does not happen. With both revision and contraction, the process by which the structure of the new belief state is elaborated is identical. First, the maximal belief commitment states are identified by taking maximal consistent subsets of the set of all beliefs in the new belief state and for each such belief theory, identifying the maximal subset of new set of disbeliefs that is consistent with it. Second, the maximal disbelief commitment states are identified by taking maximal consistent subsets of the the set of all disbeliefs in the new belief state, and for each such disbelief theory, identifying maximal subsets of the new set of beliefs that are consistent with it. An interesting possibility we have not considered here are commitment states where both the belief theory and the disbelief theory are non-maximal, yet the commitment state is maximal in the sense that no further belief or disbelief can be consistently added to it. However, this new set of possible commitment states would not make a difference to the overall dynamics of a belief state, since the class of commitment states we currently consider account for all possible beliefs and disbeliefs.

In the following subsections, we shall discuss some features of our framework in greater detail.

## Irrelevance of choice

We have seen how choice functions play a crucial role in belief change operators in the AGM framework. It could be argued that this is unintuitive, since the new belief state should be a function of the prior beliefs and the belief input, and should be independent of any preference criteria encoded into the belief change operator. If preference criteria such as the epistemic entrenchment ordering constitute part of the belief state, then they should be ultimately representable as beliefs, and should be treated as such (the formulation by Nayak *et al* of both belief state and belief input [38] as entrenchment orderings achieves this to some extent). We shall refer to this as the *principle of irrelevance of choice.* Our framework satisfies this requirement, since there is no need to select from amongst the candidate outcomes.

The irrelevance of choice is beneficial from a practical viewpoint as well. First, it permits the use of context-dependent theory preference criteria. A major concern in most studies of iterated belief change, such as the work of Nayak *et al* [37] and Darwiche and Pearl [10], has been to make the new choice function, obtained as a result of belief change, as faithful as possible to the previous one. In the AGM framework, as well as in the approach taken in [37] and [10], one would get unintuitive results if context-dependent preference criteria were to be used. In other words, it would be unintuitive to use a given entrenchment ordering at one belief change step, and a totally unrelated entrenchment ordering at the next step. This is not the case in our framework, since theory preference is orthogonal to belief maintenance. Thus, the criteria used for theory preference does not influence the outcome of a belief change step, and can be made context-dependent. The second practical benefit of the irrelevance of choice involves the use of lazy-evaluation algorithms. In the AGM framework and in systems based on this approach, all candidate outcomes must be generated in order to apply the choice function, or, alternatively, extensive computation involving entrenchment comparisons must be performed. In our framework, the computation involved in effecting belief change is minimal. A new belief is added during revision, and a new disbelief is added during contraction. Lazy evaluation techniques can be implemented in systems based on our framework, since the actual computation of commitment states can be undertaken only when required (i.e., when the agent needs to reason with, or act on its beliefs).

## Ranking belief inputs

In a recent study, Darwiche and Pearl [10] have presented a set of four postulates for iterated belief change. These postulates are motivated by the need to preserve conditional beliefs, in addition to unconditional ones. Conditional beliefs are defined via the Ramsey test [44] in the following way. Let $A \mid B$ stand for the assertion " A conditionally implies B". The conditional belief $A \mid B$ is accepted in a theory $K$ if and only if the minimal change of $K$ needed to accept $A$ also results in $B$ being accepted. In other words, $A \mid B$ is accepted in $K$ if $B$ is contained in $K^*_A$. Their postulates are as follows:

*(DP1)* If $x \models y$, then $(K^*_y)^*_x = K^*_x$.

*(DP2)* If $x \models \neg y$, then $(K^*_y)^*_x = K^*_x$.

*(DP3)* If $K^*_y \models x$, then $y \in (K^*_y)^*_x$.

*(DP4)* If $K^*_y \not\models \neg x$, then $y \notin (K^*_y)^*_x$.

Condition *(DP1)* states that if two revisions occur in sequence, with the more recent belief logically entailing the earlier one, then the first revision is redundant. We speculate that our model satisfies a reformulated version of this condition, although we shall not prove a formal result here. Condition *(DP2)* states that if two consecutive revisions contradict each other, then the more recent one prevails. This requirement is not always reasonable, in situations where belief inputs have varying degrees of credibility. For instance, the more recent belief input may have a lower level of credibility than the earlier one, in which case the reverse condition should hold. *(DP3)* requires that a belief $y$ should be retained after a more recent revision if the more recent belief, together with the current set of beliefs, entails $y$. *(DP4)* states that if a belief $y$ is not contradicted as a result of revising the current set of beliefs with $x$, then it will not be contradicted as result of revising the current set of beliefs with $y$ followed by $x$. We speculate that our approach will satisfy *(DP3)* and *(DP4)*, suitably reformulated.

Thus, condition *(DP2)* is where our approach deviates significantly from the Darwiche and Pearl approach to iterated belief change. The difference stems from different notions of success. *(DP2)* is essentially a generalization of the AGM success requirement to the iterated case. In effect *(DP2)* translates to a recency based ordering of belief inputs with more recent inputs always taking precedence over earlier ones. We differ from the approach taken by Nayak et al [37], in which a prior entrenchment ordering together with a belief input is used to generate a new entrenchment ordering, for the same reasons. Nayak et al adopt a similar recency based heuristic for ranking belief inputs, whereas we argue that recency need not be the only criterion on which belief inputs are ranked.

Our approach imposes no constraints on the rankings of belief inputs. These rankings can thus be based on the credibility of belief inputs, and can be context-dependent. In any case, such rankings are only used for theory preference and do not influence the outcome of a belief change operation in any manner. For instance, in Example 16, we get three distinct belief theories $Cn(\{b\})$, $Cn(\{b \rightarrow f\})$ and $Cn(\{b, b \rightarrow f\})$. We may select one of these based on the way we prioritize belief in $b \rightarrow f$ and its consequences, disbelief in $f$ and belief in $b$.

## Implementations

Like the AGM framework, there are considerable impediments to directly implementing our model. However, the purpose of this model is to provide a definition of competence which would serve as a starting point for the design of implementable belief change systems. We shall present the design of two such systems in the next chapter.

# Chapter 4

# Performance Models

In the previous chapter, we have presented an definition of competence for belief change under the minimal change requirement. The function of a competence theory, as we have discussed earlier, is twofold. First, it must define the best-case scenario and thus provide a yardstick by which to evaluate actual systems for belief change. Second, it must serve as a starting point for designing performance models for belief change. In this chapter, we shall demonstrate applications of our competence theory for both kinds of tasks. In the process we shall present two practically implementable systems for belief change, based on PJ-default logic.

The first system, which we shall call System *BR1*, preceded the development of the competence theory presented in the previous chapter. Several of the motivations for the competence theory first emerged while defining System *BR1*. Roughly, the belief representation scheme follows the constrained sets of theories approach, with finite sets of sentences being used in place of deductively closed theories. However, unlike the constrained sets of theories approach, System *BR1* does not follow the AGM notion of success. We evaluate System *BR1* using the AGM rationality postulates. Attempting to do this reveals a major drawback with the AGM postulates. As a result of limitations in the AGM belief representation scheme, as well as the AGM notion of success, a comparison with System *BR1* is possible only in the case when the behaviour of System *BR1* is tightly constrained. We then evaluate System *BR1* using the yardstick of the full preservation requirements. The exercise turns out to be a useful one. Since the full preservation requirements define the ideal case, evaluation of compliance with these requirements reveals the precise nature of the trade-offs made in the design of System *BR1*.

The second system, which we shall call System *BR2*, followed the development of the competence theory and was motivated by the need to demonstrate a practically implementable system based on the sets of constrained theories approach. The only trade-off made was in using representations based on finite sets of sentences instead of deductively closed theories. Predictably, the only full preservation requirements not satisfied by System *BR2* are those that insist on the use of deductively closed theories. Even so, it is possible to identify weaker versions of these conditions which reflect the shift to representations based on finite sets of sentences which are satisfied by System *BR2*.

The benefits of using the language of PJ-default logic in these two systems are twofold. First, it permits a *lazy evaluation* approach to computing belief change, in a sense that will be made clear later in this chapter. Second, a large corpus of results from research on PJ-default logic and similar formalisms can be applied in developing and implementing these systems. In the next chapter, we shall present some practical strategies for efficiently implementing systems based on PJ-default logic.

A preliminary version of part of the material on System *BR1* has appeared earlier in [24] and [25].

## 4.1 PJ-Default Logic

PJ-default logic is a variant of default logic in which default rules are restricted to be prerequisite-free and semi-normal (i.e., a PJ-default rule is of the form $\frac{:\beta}{\gamma}$ such that $\beta \models \gamma$). PJ-default logic improves over Reiter's default logic [46] by avoiding cases where Reiter's logic is too weak, preventing the derivation of "reasonable" conclusions (such as in the disjunctive default problem) as well as cases where Reiter's logic is too strong, permitting the derivation of unwanted conclusions (for a detailed discussion of these issues, see [14]). This approach has other useful properties as well, such as semi-monotonicity, the guaranteed existence of extensions, weak orthogonality of extensions and a constructive definition for extensions. PJ-default extensions are defined as follows:

**Definition 3** [14] *Let (W,D) be a prerequisite-free semi-normal default theory. Define:*

$E_0 = (E_{J_0}, E_{T_0}) = (Cn(W), Cn(W))$

$E_{i+1} = (E_{J_{i+1}}, E_{T_{i+1}}) = (Cn(E_{J_i} \cup \{\beta \wedge \gamma\}), Cn(E_{T_i} \cup \{\beta\}))$

*where*

$i \geq 0,$

$\frac{:(\beta \wedge \gamma)}{\beta} \in D,$

$\neg(\beta \wedge \gamma) \notin E_{J_i}.$

*Then E is a PJ-extension for (W,D) iff*

$E = (E_J, E_T) = (\bigcup_{i=0}^{\infty} E_{J_i}, \bigcup_{i=0}^{\infty} E_{T_i}).$

In the rest of the paper, whenever we refer to an extension, we shall refer to the $E_T$ part of a PJ-extension.

Ghose and Goebel [23] have earlier defined a belief change framework in which a belief state is represented as a potentially inconsistent set of sentences, together with a partial order on these sentences. An operator is defined that identifies consistent subsets of this set of sentences, that respect the partial order as well as some set of disbelief constraints. A translation from PJ-default theories to this framework is defined such that the process of identifying PJ-default extensions is shown to be equivalent to the process of identifying consistent subsets of sentences using the operator mentioned above, with a partial order which assigns higher priority to sentences obtained from $W$ (given a PJ-default theory $(W, D)$) than sentences obtained from $D$ and with the set of disbelief constraints consisting of the conjunction of the negations of the justifications of each PJ-default rule.

## 4.2 *BR1*: Definition

System *BR1* was originally motivated by the following observations:

- A belief state is best represented as a collection of theories. Given that *minimal change* is a guiding principle in belief revision, it could be argued that instead of selecting some outcomes of the belief change and discarding the others (thus losing potentially useful beliefs), all outcomes should be retained, provided that there exists a compact and elegant way of representing these multiple possible outcomes. Such representation languages exist; nonmonotonic formalisms are immediate candidates for compactly representing the possibly inconsistent and incomplete picture of the world that each belief state corresponds to. The observation that real-life agents typically have incomplete (and sometimes inconsistent) knowledge about the world is independent justification for choosing such an approach.

- In order that previous contractions are taken into consideration in determining the outcome of a belief change step in the same way as previous revision steps, it is desirable to maintain a record of both revisions as well as contractions. We achieve this by maintaining a set of *belief constraints* (these may be of two kinds: constraints specifying beliefs that must necessarily be held, and constraints specifying beliefs that must necessarily not be held) which may be viewed as the integrity constraints of a belief system; every theory constituting a belief state must respect them.

- Beliefs originally held to be true can become tentative as a result of belief change if they are contradicted or brought into question (the intuition is that a belief becomes questionable if it is not in every possible outcome of the belief change step) by the new evidence. In syntactically-oriented nonmonotonic formalisms, this can be viewed as a process of demotion from a fact to a default.

The belief representation scheme in System *BR1* is similar to the scheme in the constrained sets of theories approach described in the previous chapter. Thus, a belief state consists of a set of theories denoting beliefs, together with a theory denoting the set of disbeliefs. In addition, a belief state consists of the following three elements:

- A set of necessary belief constraints.
- A set of necessary disbelief constraints.
- A constraint prioritization relation.

Let $\{K_1, \ldots, K_n\}$ be the set of belief theories in a given belief state. The *necessary belief set* $BC_{belief}$, for a given belief state, is a set of sentences such that, for every $x \in BC_{belief}$ and for all $K_i$ in that state, $K_i \models x$. The *necessary disbelief set* $BC_{disbelief}$, for a given belief state, is a set of sentences such that, for every $\neg x \in BC_{disbelief}$ and for all $K_i$ in that state, $K_i \not\models x$. Notice that $BC_{disbelief}$ contains the *negations* of beliefs that may not be held. A *tentative belief* is a belief that is an element of some, but not all $K_i$ in that belief state. We define the *belief constraint set* $BC = BC_{belief} \cup BC_{disbelief}$. Updating the belief constraint set requires identifying maximal parts of a constraint that is consistent with some others. We shall not present a formal notion of a part of constraint here, but the following example should make our intuitions clear.

**Example 26** Consider retracting both $a$ and $b$ in a single step from a belief state. This corresponds to ensuring that the formula $a \vee b$ is not believed in this state. We shall therefore add a necessary disbelief constraint $\neg a \wedge \neg b$. Let the next belief change step be a revision with $a$, which corresponds to adding a necessary belief constraint $a$. Clearly, it is impossible to enforce disbelief in both $a$ and $b$, and belief in $a$ at the same time. The two constraints cannot be satisfied at the same time. One option is to delete the disbelief constraint $\neg a \wedge \neg b$ since it contradicts the newly acquired constraint. However, if we do so, we shall allowing the belief $b$ to be held. All that we have been told since both $a$ and $b$ were retracted is that $a$ is to be believed again. It makes better sense to retain the requirement for disbelief in $b$. Thus, we should retain the maximal part of the disbelief constraint $\neg a \wedge \neg b$ that is consistent with the new belief constraint $a$, i.e., the disbelief constraint $\neg b$. □

A set-theoretic representation of each constraint such that each conjunct in a constraint is clearly identifiable as a set element would enable us to identify maximal subsets that are consistent with other constraints. Representing belief constraints in clausal form satisfies these requirements. To be able to distinguish between the elements of $BC_{belief}$ and $BC_{disbelief}$, we adopt a syntactic convention such that every disbelief constraint $\phi$ is written as $-\phi$. Thus, constraints requiring necessary belief in $a$ and $a \rightarrow b$ will be written as $\{\{a\}\}$ and $\{\{\neg a, b\}\}$ respectively, while a constraint requiring disbelief in $((c \vee d) \wedge e)$ will be written as $-\{\{\neg c, \neg e\}, \{\neg d, \neg e\}\}$. Belief change will be viewed as the process of adding a new belief constraint (in the case of revision, a necessary belief constraint, and in the case of contraction, a necessary disbelief constraint).

We require that the belief constraint set be totally ordered; we shall refer to this total order $\prec$ as the *constraint prioritization* (we shall write $x_j \prec x_i$ if constraint $x_i$ has a higher priority than constraint $x_j$). This is similar to the orderings used in a variety of belief change frameworks, including the AGM epistemic entrenchment ordering, but there are significant differences. Whereas epistemic entrenchment requires that all beliefs be prioritized, we require that only the current set of belief constraints be prioritized. We shall see later that the size of the belief constraint set can typically be expected to be much smaller than the size of the theories constituting a belief state. As well, the size of the belief constraint set does not grow with time and may shrink as belief constraints are discarded (as a result, for example, of newer constraints contradicting them). This also represents a more principled approach to prioritizing beliefs, since it does not suffer from the contradictions in belief prioritization in the AGM framework pointed out earlier. Notice also, that unlike AGM epistemic entrenchment, this ordering does not uniquely determine which subset of the possible outcomes is finally selected. It only determines what the new set of belief constraints should be. To draw a database analogy, the constraint prioritization ranks only the current integrity constraints, whereas epistemic entrenchment requires that every assertion in the database be ranked. Several obvious heuristics suggest themselves as candidate constraint prioritization policies in the absence of any other information on the relative reliability of the belief constraints. In the case of *revision* (in the sense of Katsuno and Mendelzon [29], i.e., for belief change in static worlds), a constraint prioritization base on the recency of the belief inputs appears to be appropriate. In the case of *update* [29](i.e., belief change in dynamic worlds), one might choose to accord the highest priority to physical laws at all times, since these are never questioned or discarded.

System *BR1* uses a PJ-default theory, together with a constraint prioritization relation to represent a belief state. The process of belief change involves two steps. First, the belief constraint set is updated. Based on the updated belief constraint set, the PJ-default theory representing the previous belief state is modified.

**Definition 4** *A belief constraint $\Omega$ (where $\Omega = \phi$ if $\Omega$ requires necessary belief in $\phi$ and $\Omega = \neg\phi$ if $\Omega$ requires necessary disbelief in $\phi$) is said to be* **compatible** *with a set of belief constraints BC (we write $\Omega \cup BC$ is compatible) if and only if for all $x_i \in BC_{belief}$ and all $\neg y_j \in BC_{disbelief}$, $\Omega \wedge (\bigwedge_i x_i) \wedge (\bigwedge_j \neg y_j)$ is satisfiable.*

We must be able to identify subsets of an individual belief constraint (viewing each constraint as a set of clauses) that are compatible with a set of belief constraints. The operator $\uparrow$ that identifies such subsets is defined below.

**Definition 5** *Let bc be a belief constraint and BC be a set of belief constraints. Then*

$$bc \uparrow BC = \{x \subseteq bc \mid \ (x \cup BC \text{ is compatible}) \wedge$$
$$(\forall x' \text{ such that } x \subset x' \subseteq bc \text{ , } x' \cup BC \text{ is incompatible}\}$$

Updating the set of belief constraints involves starting with the constraint of highest priority and working downwards, adding as many constraints (or parts of constraints) of lower priority as can be *compatibly* added. In the definition below, we assume that $BC_{old}$ is the old set of constraints, $\Omega$ is the new constraint being added in the current belief change step and $BC_{inter} = BC_{old} \cup \Omega$. We assume that the constraint prioritization relation $\prec$ is updated to reflect the ranking of $\Omega$ relative to the other constraints. As well, every $bc_i$ belongs to $BC_{inter}$ and $bc_j \prec bc_k \leftrightarrow k < j$ (i.e. constraints of a higher priority have a smaller subscript).

$$\begin{aligned} BC_{new} = \{Y \subseteq BC_{inter} \mid\ & Y = \textstyle\bigcup_{i\geq 1} Y_i, \\ & \forall i \geq 1(((Y_i = \{bc_i\}) \wedge (bc_i \cup (\textstyle\bigcup_{j=1}^{i-1} Y_j) \text{ is compatible})) \\ & \vee((Y_i = \textstyle\bigcap(bc_i \uparrow (\bigcup_{j=1}^{i-1} Y_j \cup \Theta)) \wedge (bc_i \cup (\textstyle\bigcup_{j=1}^{i-1} Y_j) \text{ is incompatible})) \end{aligned}$$

It is not necessary to define the new constraint prioritization relation. This can be generated dynamically, driven by the context. The only requirement is that a constraint prioritization relation exist at every belief change step.

In the definition above, notice that in the process of collecting constraints from highest to lowest priority, if a constraint turns out to be incompatible with the set of higher priority constraints collected so far, we identify maximal subsets of the constraint (given the representation of a constraint as a set of clauses) that are compatible with the collected constraints, and add the intersection of these subsets to the set of collected constraints. The following examples clarify this.

**Example 27** Let $BC_{old} = \{\{\{\neg a, b\}\}, \{\{\neg b\}\}\}$ and let $\prec$ be such that $\{\{\neg b\}\} \prec \{\{\neg a, b\}\}$. Let $\Omega = \{\{a\}\}$. Thus $BC_{inter} = \{\{\{\neg a, b\}\}, \{\{\neg b\}\}, \{\{a\}\}\}$ and the updated prioritization relation $\prec$ is such that $\{\{\neg b\}\} \prec \{\{\neg a, b\}\} \prec \{\{a\}\}$. Then $BC_{new} = \{\{\{a\}\}, \{\{\neg a, b\}\}\}$. □

**Example 28** Let the initial belief state have a single constraint requiring disbelief in $a \vee b$. Thus $BC_{old} = \{-\{\{\neg a\}, \{\neg b\}\}\}$ and the relation $\prec$ is empty. Let $\Omega = \{\{a\}\}$ to be added at the highest priority level. $BC_{inter} = \{\{\{a\}\}, -\{\{\neg a\}, \{\neg b\}\}\}$ and the updated $\prec$ relation is such that $-\{\{\neg a\}, \{\neg b\}\} \prec \{\{a\}\}$. Then $BC_{new} = \{-\{\{\neg b\}\}, \{\{a\}\}\}$ and the resulting $\prec$ relation is such that $-\{\{\neg b\}\} \prec \{\{a\}\}$. Thus, while a constraint requiring disbelief in both $a$ and $b$ is not compatible with a necessary belief constraint in $a$, a subset which requires disbelief in $b$ only is compatible.□

Using PJ-default logic as the belief representation language enables us to represent belief constraints in a PJ-default theory. The set of facts $W$ of a PJ-default theory corresponds to the beliefs that the agent is constrained to necessarily hold, since $W$ will be contained in every consistent belief set corresponding to that belief state (i.e., every PJ-extension). Since every PJ-default rule is of the form $\frac{:\beta\wedge\gamma}{\beta}$, $\neg\gamma$ corresponds to the theory that the agent is constrained to necessarily disbelieve. The process of demoting a belief that has been contradicted, or made questionable (meaning that it is no longer in every possible outcome of that belief change step), as a result of a belief change step to the status of a tentative belief involves removing a formula from $W$ and adding a new PJ-default rule containing this formula as its consequent to $D$. Discredited beliefs are thus never totally discarded in our framework in anticipation of future situations in which these beliefs could be consistently held again. Belief change in our framework thus involves mapping one PJ-default theory to another. The possible consistent belief sets that may be held in a given belief state corresponds to the extensions of the PJ-default theory representing that belief state. The process is graphically illustrated in Figure 1.

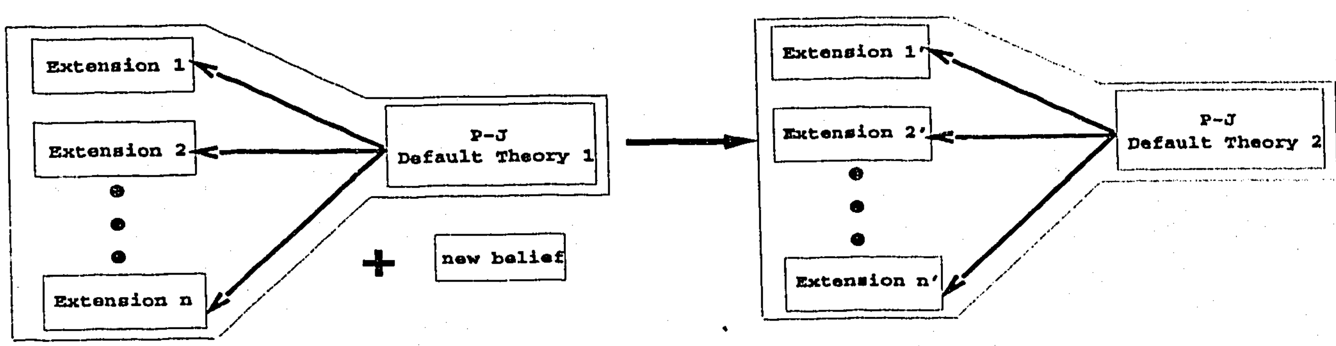


Figure 4.1: Belief State Transformation

In the rest of this chapter, we shall refer to the $E_T$ component of a PJ-extension $(E_T, E_J)$ as an *extension*.

The new PJ-default theory (W',D') is obtained from the previous belief state $(W, D)$ and the new belief constraint set $BC_{new}$ as follows:

$$
\begin{aligned}
W' &= BC_{Belief_{new}} \\
D' &= \{\frac{: \delta_i \wedge (\bigwedge BC_{Disbelief_{new}})}{\delta_i} \mid \delta_i \in (W'' - W') \} \cup \\
& \quad \{\frac{: \beta_i \wedge (\bigwedge BC_{Disbelief_{new}})}{\beta_i} \mid \frac{: \beta_i \wedge \phi_i}{\beta_i} \in D \}
\end{aligned}
$$

where $W'' = W$ in case the operation is contraction and $W'' = W \cup \{x\}$ where $x$ is the new necessary belief constraint in case the operation is revision. Here $\bigwedge BC_{Disbelief_{new}}$ stands for the conjunction of all the elements of $BC_{Disbelief_{new}}$. Thus, if $BC_{Disbelief_{new}} = \{-\{\{\neg b\}\}, -\{\{\neg a\}\}\}$, then $\bigwedge BC_{Disbelief_{new}} = \neg b \wedge \neg a$.

Notice that the justification for every default rule in PJ-default theories representing belief states in our framework is identical and corresponds to the set of necessary disbelief constraints. We define the notion of *uniform* default theories as follows:

**Definition 6** *A semi-normal default theory* $(W, D)$ *is said to be uniform if for any two default rules* $\frac{\alpha_i:\beta_i \wedge \gamma_i}{\beta_i}, \frac{\alpha_j:\beta_j \wedge \gamma_j}{\beta_j} \in D, \gamma_i = \gamma_j$ *(if* $\mid D \mid= 1$*, then the theory is trivially uniform).*

Clearly, every PJ-default theory representing a belief state in our framework is *uniform.*

We require that the dummy default rule $\frac{:\top}{\top}$ be an element of $D$ for every $(W, D)$ representing a belief state. This is to enable us to record necessary disbelief constraints even if there are no tentative beliefs.

**Example 29** Let the initial belief state be given by the uniform PJ-default theory $(W, D)$ where $W = \{a, a \rightarrow b, c, c \rightarrow d\}$ and $D = \{\frac{:\top}{\top}\}$. Notice that there are no necessary disbelief constraints at this point. Assume that the elements of $W$ were obtained in a single belief change step. Thus, there is a single necessary belief constraint and $BC_{Belief_{old}} = \{\{\{a\}, \{\neg a, b\}, \{c\}, \{\neg c, d\}\}\}$.

Let us now retract the belief $b$.

- **Belief Constraints Update** Let $\{\{a\}, \{\neg a, b\}, \{c\}, \{\neg c, d\}\} \prec -\{\{\neg b\}\}$.

$$\begin{aligned} BC_{Belief_{new}} &= \{\{\{c\}, \{\neg c, d\}\}\} \\ BC_{Disbelief_{new}} &= \{\{-\{b\}\}\} \end{aligned}$$

- **Default Theory Update**

$$\begin{aligned} W' &= \{c, c \rightarrow d\} \\ D' &= \{\frac{: a \wedge \Gamma}{a}; \frac{: (a \rightarrow b) \wedge \Gamma}{(a \rightarrow b)}; \frac{: \top \wedge \Gamma}{\top}\}, \ \textit{where} \ \Gamma = \neg b \end{aligned}$$

Notice that, as a result of the contraction, $a$ and $(a \rightarrow b)$ become *tentative beliefs*. Notice also that there are two possible belief sets that may be consistently held in the final belief state (corresponding to the two extensions of $(W', D')$) given by $Cn(\{c, c \rightarrow d, a\})$ and $Cn(\{c, c \rightarrow d, a \rightarrow b\})$ and that the belief $b$ is not held in either of these. The new constraint prioritization is $\{\{c\}, \{\neg c, d\}\} \prec -\{\{\neg b\}\}$.□

By representing each belief state as a uniform PJ-default theory, we have factored out the question of which theory (extension) to choose as the currently operative set of beliefs from the process of belief change. The user, or agent, could thus employ a variety of techniques to actually pick the currently operative set of beliefs. If priorities on beliefs are used to select the currently operative set of beliefs (as in AGM epistemic entrenchment), then our framework would permit dynamic prioritization of beliefs. In other words, different orderings could be used at different times to select theories without causing inconsistencies or unreasonable outcomes from the belief change process. The more crucial priority relation, however, is the one that determines the nature of the new belief state. In our case, this is the *constraint prioritization* relation. Unlike the AGM epistemic entrenchment relation and Nebel's epistemic relevance ordering (which, like our constraint prioritization relation, determine the nature of the new belief state),we provide a clear prescription of how the relation evolves over iterated belief change steps and how the new evidence is integrated into this relation. An additional advantage with our framework is that we permit, in addition to the traditional operations of expansion, revision and contraction, the "undoing" of a contraction step; it simply involves the removal of the relevant necessary disbelief constraint.

## 4.3 *BR1*: Relation to the AGM framework

Our formalization cannot be evaluated using the AGM postulates directly, for the following two reasons. First, the AGM postulates consider transitions between belief states represented as a single deductively closed propositional theory. Our operator maps between belief states

represented as collections of theories (the multiple possible extensions of the PJ-default theories). Second, since the AGM postulates consider belief change as a single step process, it is difficult to evaluate "rationality" over iterated belief change steps. It is possible, however, to articulate a reformulated version of these postulates, and show that our framework satisfies them under certain conditions. We shall formalize these conditions first.

**Definition 7** *Let a belief change operation result in the introduction of a new constraint $x$ in a belief state with a constraint set denoted by $BC$. Let $y \in BC$ be such that $y$ is incompatible with $x$ and there is no $z \in BC$ such that $y \prec z$ and $z$ is incompatible with $x$. The belief change operation is* imperative *iff $y \prec x$.*

Thus, an imperative belief change operation introduces a constraint that has a higher priority than the existing constraint of the highest priority that is incompatible with it. Since our framework is general enough to permit any prioritization of the belief constraints, it is possible for a belief change operation to fail (in the case that there exists a belief constraint, with higher priority than the newly introduced constraint, which is incompatible with the newly introduced constraint. If a belief change operation is imperative, the operation is guaranteed to succeed. Since every belief change step in the AGM framework must succeed, our framework satisfies the AGM postulates only in the case of imperative belief change operations.

The second condition involves the prevention of beliefs that were previously suppressed by the existence of a disbelief constraint from reappearing in a belief state when this disbelief constraint is discarded.

**Example 30** Consider a belief state given by $W = \{\}$ and $D = \{\frac{:\neg a \wedge a \wedge (a \rightarrow b)}{\neg a}, \frac{:\top \wedge a \wedge (a \rightarrow b)}{\top}\}$. There is a single, empty, belief set corresponding to $Cn(\top)$, which is the only extension of $(W, D)$. Let the belief constraint set in this belief state consist of a single disbelief constraint $-\{\{a\}, \{\neg a, b\}\}$. Let the belief $b$ be retracted from this belief state. We shall get a new belief constraint set given by a single disbelief constraint $-\{\{\neg b\}\}$. The new belief state will be given by $W' = \{\}$ and $D' = \{\frac{:\neg a \wedge \neg b}{\neg a}, \frac{:\top \wedge \neg b}{\top}\}$. This default theory has one extension, $Cn(\neg a)$. Thus we get a belief state with a belief set containing $\neg a$ as result of contracting $b$ from a belief state containing a single, empty, belief set. This clearly violates the AGM contraction postulate which requires that the contracted belief set should be a subset of the original belief set, yet the behaviour is perfectly rational. The tentative belief $\neg a$ reappears in a belief set as a consequence of the removal of the disbelief constraint that caused this tentative belief to be suppressed.□

Clearly, only operations which do not display such behaviour can be related to the AGM framework.

**Definition 8** *Let a belief change operation introduce a belief constraint $x$ in a belief state $(W, D)$ with a belief constraint set $BC$. The operation is* stable *iff there exists no disbelief constraint $y \in BC$ .s.t $y \prec x$ and there exists $\frac{:\beta \wedge \gamma}{\beta} \in D$ where $y \cup \beta$ is incompatible.*

Before we state and prove the results relating our framework to the AGM postulates, we shall establish a connection between THEORIST system developed by Poole, Goebel and Aleliunas [42] and uniform PJ-default theories that simplifies the proofs. The THEORIST framework envisages a knowledge base comprising of a set of closed formulas that are necessarily true, called *facts*, and a set of possibly open formulas that are tentatively true, called *hypotheses*. Default reasoning in this framework involves identifying *maximal scenarios* (or extensions), where a scenario consists of the set of facts together with some subset of the set of ground instances of the hypotheses which is consistent with the set of facts. The framework can be augmented with *constraints*, which are closed formulas such that every THEORIST scenario is required to be consistent with the set of constraints. Following [42], we can present the following definition of a maximal scenario.

**Definition 9** *For a THEORIST system* $(F, H, C)$ *where* $F$ *is the set of facts,* $H$ *is the set of hypotheses and* $C$ *is the set of constraints, such that every element of* $F$, $H$ *and* $C$ *is a ground formula, a* maximal scenario *is a set* $F \cup h$ *such that* $h \subseteq H$ *and* $F \cup h \cup C$ *is satisfiable.*

We shall present a translation which is simpler than the one presented in [14] because we are dealing with uniform PJ-default theories rather than general ones. In the following, $S(F, H, C)$ refers to the set of maximal scenarios of the THEORIST system $(F, H, C)$. As well, $E(\Delta)$ refers to the set of extensions of the default theory $\Delta$.

**Definition 10** *Let* $(F_{(W,D)}, H_{(W,D)}, C_{(W,D)})$ *denote the THEORIST-translation of the uniform PJ-default theory* $(W, D)$. *Then:*

$$F_{(W,D)} = W$$
$$H_{(W,D)} = \{\beta \mid \tfrac{:\beta\wedge\gamma}{\beta} \in D\}$$
$$C_{(W,D)} = \{\gamma \mid \tfrac{:\beta\wedge\gamma}{\beta} \in D\}$$

**Theorem 14** *Let* $(F_{(W,D)}, H_{(W,D)}, C_{(W,D)})$ *denote the THEORIST-translation of the uniform PJ-default theory* $(W, D)$. *Then* $S(F_{(W,D)}, H_{(W,D)}, C_{(W,D)}) = E((W, D))$

**Proof:** The proof follows directly from the equivalence of the definitions of THEORIST maximal scenario computation and PJ-default extension computation, given that $E_{J_i} = C_{(W,D)}$ at every step in the computation of extensions for a uniform PJ-default theory $(W, D)$. □

Thus, the facts $F$ correspond to the necessary belief constraints. The constraints $C$ correspond to the necessary disbelief constraints in our system. The hypotheses

$H$ correspond to the tentative beliefs. The process of belief change can thus be equivalently defined as computing the new set of belief constraints, replacing $F$ and $C$ accordingly and demoting beliefs from $F$ to $H$ if necessary.

We shall interpret postulate (1-), as one of way of articulating the following *principle of categorical matching* stated by Gärdenfors and Rott in [22] which requires that the representation of a belief state after a belief change has taken place should be of the same format as the representation of the belief state before the change. For postulates (2-) through (8-), we reformulate every condition on knowledge sets to apply to every extension of the PJ-default theory representing a belief state. For postulates (7-) and (8-) we can actually prove a stronger condition in the case that the antecedent in (8-) is satisfied. If the antecedent is not satisfied, there appears to be no obvious way to reformulate postulate (7-).

**Theorem 15** *For imperative and stable operations, the contraction operator $-$ satisfies:*

1. *The principle of categorical matching.*
2. $\forall e' \in E((W,D)^-_A)$, *there exists some $e$ where $e \in E((W,D))$ s.t. $e' \subseteq e$.*
3. *If* $\forall e : (e \in E((W,D))) \supset (e \not\models A)$, *then* $E((W,D)^-_A) = E((W,D))$.
4. *If* $\not\models A$, *then* $\forall e : (e \in E(W,D)^-_A) \supset (e \not\models A)$.
5. *If* $\forall e' : (e' \in E((W,D))) \supset (e' \models A)$ *then for every* $e' \in E((W,D))$, *there exists some $e$ where* $e \in E(((W,D)^-_A)^+_A)$ *s.t* $e' \subseteq e$.
6. *If* $\models A \leftrightarrow B$ *then* $E((W,D)^-_A) = E((W,D)^-_B)$.
7. *If* $\forall e : (e \in E((W,D)^-_{A\wedge B})) \supset (e \not\models A)$ *then* $E((W,D)^-_{A\wedge B}) = E((W,D)^-_A)$.

**Proof:** We shall prove these results using the THEORIST-translations $(F,H,C)$ of $(W,D)$.

1. Obvious.

2. To prove: $\forall e' \in S((F,H,C)_{\overline{A}})$, there exists some e where $e \in S((F,H,C))$ s.t. $e' \subseteq e$.

   Let $(F,H,C)_{\overline{A}} = (F',H',C')$. Assume the contrary. Thus, there must exist some $e_x \in S((F',H',C'))$ s.t. there exists no $e \in S((F,H,C))$ where $e_x \subseteq e$. Let $e_x = Cn(F' \cup h')$ where $h' \subseteq H'$. Two cases are possible:

   (a) $F = F'$. In this case, $H = H'$. Our assumption holds iff $e_x \cup C$ is unsatisfiable. By stability, this is impossible. Thus, our assumption does not hold.

   (b) $F' \subset F$. Thus $H \subset H'$. Let $y = H' - H = F - F'$. If $h' \subseteq H$, then $e_x$ is included in some $e \in S((F,H,C))$, since $F'$ is included in $F$, $h'$ is included in $H$, $Cn(F' \cup h')$ is satisfiable (by virtue of being a scenario of $(F',H',C')$) and $Cn(F' \cup h')$ is consistent with $C$, as a consequence of consistency with $C'$ and stability. If $h' \not\subseteq H$, then $y'' \subseteq y$, where $y'' = h' - H$. Thus, $Cn(F' \cup h') = Cn(F' \cup y'' \cup h'')$ where $h'' \subseteq H$. Then $e_x$ is included in some $e \in S((F,H,C))$ since $F' \cup y''$ is included in $F$, $h''$ is included in $H$, $Cn(F' \cup y'' \cup h'')$ is satisfiable (by virtue of being a scenario of $(F',H',C')$), and $Cn(F' \cup y'' \cup h'')$ is consistent with $C$ as a consequence of consistency with $C'$ and stability.

3. To prove: If $\forall e : (e \in S((F,H,C))) \supset (e \not\models A)$, then $S((F,H,C)_{\overline{A}}) = S((F,H,C))$.

   Let $(F,H,C)_{\overline{A}} = (F',H',C')$. Clearly $F \not\models A$. Thus, $F = F'$, $H = H'$. Then, by stability, since all elements of $S((F,H,C))$ are consistent with $C$, they will be consistent with $C'$ too.

4. To prove: If $\not\models A$, then $\forall e : (e \in S((F,H,C)_{\overline{A}}) \supset (e \not\models A)$.

   $C \models \neg A$. Hence proved.

5. To prove: If $\forall e' : (e' \in S((F,H,C))) \supset (e' \models A)$ then for every $e' \in S((F,H,C))$, there exists some $e$ where $e \in S(((F,H,C)^-_A)^+_A)$ s.t. $e' \subseteq e$.

   Let $((F,H,C)^-_A)^+_A = (F',H',C')$. Let $e_x = Cn(F \cup h)$, where $h \subseteq H$ be some arbitrarily chosen element of $S((F,H,C))$. Since $H \subseteq H'$, $h \subseteq H'$. Two cases are possible:

   (a) $F \models A$. Then $F' \subseteq F$. Let $y = F - F'$. Clearly $y \subseteq H'$. Thus, $Cn(F \cup h) = Cn(F' \cup y \cup h)$. $Cn(F' \cup y \cup h)$ is satisfiable by virtue of being a scenario of $(F,H,C)$. Clearly $Cn(F' \cup y \cup h)$ includes $F'$. As well, $y \cup h \subseteq H'$. Since $e_x$ is consistent with $C$, $Cn(F' \cup y \cup h)$ is consistent with $C'$, by stability. Thus $e_x$ must be included in some $e \in S((F',H',C'))$.

   (b) $F \not\models A$. Then $F \subset F'$. Clearly $F' - F = A$. Since $Cn(F \cup h) \models A$, $Cn(F \cup h) = Cn(F \cup A \cup h) = Cn(F' \cup h)$. Clearly $Cn(F \cup A \cup h)$ includes $F'$. As well, $h \subseteq H'$. $Cn(F \cup A \cup h)$ is satisfiable by virtue of being a scenario of $(F,H,C)$. Since $e_x$ is consistent with $C$, $Cn(F \cup A \cup h)$ is consistent with $C'$, by stability. Thus, $e_x$ must be included in some $e \in S((F',H',C'))$.

6. To prove: If $\models A \leftrightarrow B$ then $S((F,H,C)^-_A) = S((F,H,C)^-_B)$.

   Trivially true.

7. To prove: If $\forall e : (e \in S((F,H,C)^-_{A \wedge B})) \supset (e \not\models A)$then $S((F,H,C)^-_{A \wedge B}) = S((F,H,C)^-_A)$.

   Let $(F,H,C)^-_A = (F',H',C')$ and $(F,H,C)^-_{A \wedge B} = (F'',H'',C'')$. By stability, and by the precondition, it is easy to see that every element of $S((F',H',C'))$ is consistent with $C''$ and every element of $S((F'',H'',C''))$ is consistent with $C'$. The precondition implies that $F'' \not\models A$. Since $F'' \subseteq F$ and $F'' \not\models A$, $F'' \subseteq F'$. Assume that $F'' \subset F'$. Then there must exist some $x \in F'$ such that $F'' \cup \{x\} \models A \wedge B$. But this is impossible since $F'' \cup \{x\} \subseteq F'$. Hence $F'' = F'$. Thus $H' = H''$. Hence proved.

Over a single step, and starting with a deductively closed theory, our framework is identical ot the AGM framework in the sense that the outcomes generated by our framework are identical to the choices available to the AGM *selection function.*

**Theorem 16** *For a uniform PJ-default theory* $(W, D)$, $E((W,D)^{-}_{A}) = K \downarrow A$ *if* $W = K$ *where* $K$ *is a belief set,* $D = \{\frac{:\top}{\top}\}$, *the contraction operation is imperative and the initial constraint prioritization relation is empty.*

**Proof:** The proof follows directly from the definitions of the removal operation $\downarrow$, the constraint update operation and imperative belief change operations.□

Thus, if we were to start with a deductively closed theory as the set of facts, and an empty set of defaults, then the set of extensions of the default theory obtained after contracting a belief $A$, would correspond precisely to the set of possible outcomes that the selection function in partial meet contraction. Whereas partial meet contraction requires that a choice is actually made, we do not require any choices, but retain all the multiple possible outcomes compactly represented as a PJ-default theory.

## 4.4 *BR1*: Other Related Work

The following theorem shows how our approach relates to Nebel's base contraction operator [39].

**Theorem 17** *Let* $(W, D)$ *be a uniform PJ-default theory with* $W = B$, *where* $B$ *is a finite belief base, and* $D = \{\frac{:\top}{\top}\}$. *Then* $Cn(B \simeq A) = Cn((\bigvee E((W,D)^{-}_{A})) \wedge (B \vee \neg A))$ *if the contraction operation is imperative and the initial constraint prioritization relation is empty.*

**Proof:** The proof follows directly from the definitions of the operation $\simeq$, the constraint update operation and imperative belief change operations.

As with our framework, a belief that becomes suppressed as a result of a contraction operation can be recovered in Nebel's framework when the belief state is revised with the contracted belief. However, our framework permits an explicit operation to undo a contraction, which can also result in beliefs being recovered. Such an operation is not possible in Nebel's framework.

With their commitment to producing a unique outcome for the belief change operation, both the AGM and Nebel frameworks render too many potentially useful beliefs unusable (notice that they are not actually discarded, but can be recovered later under certain circumstances); in the AGM framework, this is a consequence of taking the intersection of the selected outcomes, while in Nebel's case, this is a consequence of taking the disjunction of every possible outcome. Our framework retains every possible outcome at all times, and thus does not suffer from this problem.

Brewka [6] shows how belief revision can be viewed as a simple process of adding new information to theories represented in his *preferred subtheories framework*, which is a generalization of the THEORIST framework [42]. Brewka's framework of preferred subtheories differs from THEORIST in two significant ways. First, *facts* are done away with, making every formula in the knowledge base potentially refutable. Second, one is allowed to define a partial order on the formulas in the knowledge base. A preferred subtheory, Brewka's analogue of a THEORIST maximal scenario, is a consistent subset of the knowledge base constructed by starting with formulas with the highest priority (as defined by the partial order) and progressively adding as many formulas of lower priority levels as can be consistently added. As with THEORIST, a knowledge base can have several preferred subtheories. Brewka shows in [6] that a knowledge base of this kind can be revised by simply adding the new formula and augmenting the partial order to incorporate any ordering relationships that might exist between this formula and the existing elements of the knowledge base. Also, if this framework is augmented to include THEORIST-style constraints, and a partial order is defined on the set containing both the formulas representing hypotheses and formulas representing constraints, then contraction is shown to be a simple case of adding a constraint to the knowledge base and augmenting the partial order. The improvements achieved by Brewka's belief change framework over earlier ones are twofold. First, the belief change operator is simple and totally incremental. Second, earlier information is not thrown away, but is retained in an elegant fashion. Nebel [40] establishes a restricted form of equivalence between nonmonotonic inference and belief change along similar lines.

In the case that the new belief (either a new hypotheses, as in revision, or a new constraint, as in contraction) always has a higher priority, under the partial ordering, than all existing beliefs, Brewka's framework turns out to be very similar to ours. As the following example shows, his framework avoids the problem of spurious beliefs in most cases.

**Example 31** Let the initial knowledge base consist of the set $\{a, a \rightarrow b\}$ of hypotheses with no ordering relationship being defined on the hypotheses. In order to contract $b$ from this knowledge base, we add the constraint $\neg b$, written as $< \neg b >$ to the knowledge base, together with the ordering relations $< \neg b > \geq a$ and $< \neg b > \geq a \rightarrow b$. We get two maximal scenarios , one containing $a$ and the other containing $a \rightarrow b$. Further revision of the knowledge base with $a$ results in the addition of this hypotheses at a higher priority level than all existing elements (hypotheses or constraints) of the knowledge base. There is only one maximal scenario at this point, consisting of $a$ and its logical consequences. □

The similarity of Brewka's framework to ours is not surprising, given that we use, like Brewka, nonmonotonic theories which can generate possibly many different consistent sets of beliefs, to represent a belief state. Like Brewka, our approach is incremental, and information is never thrown away. Our choice of nonmonotonic formalism is very similar too, given the results in [14] relating PJ-default logic to THEORIST with constraints. However, since Brewka does not explicitly account for the maintenance of belief constraints, his formalization may provide unintuitive results as the following example shows.

**Example 32** Consider an initial knowledge base containing only one hypotheses and no constraints $\{a\}$. Let us now contract $a \vee b$ from this knowledge base. This entails the addition of the constraint $< \neg(a \vee b) >$ to the knowledge base, and augmenting the partial order such that the new constraint has higher priority than all existing elements of the knowledge base. If one were to revise the knowledge base with $b$, there would be one maximal scenario containing both $a$ and $b$. Notice, however, that the new evidence obtained since retracting $a \vee b$ from the knowledge base does not warrant renewed belief in $a$. The problem arises because the presence of $b$ at a higher priority level disables the constraint $< \neg(a \vee b) >$.□

We improve upon Brewka's work by explicitly accounting for the maintenance of belief constraints. Necessary disbelief constraints are treated as a set of formulas to be explicitly disbelieved. We update this set at every belief change step, by retaining as many constraints, or parts of constraints, as are compatible with more recent constraints. Thus in the previous example, we would update this theory to account for revision with $b$ by removing $b$ from the set of necessary disbelief constraints, but leaving $a$ intact. The problems with Brewka's framework stems from the fact that it uses syntactic units (the constraints) which are enabled or disabled as whole units and not in terms of the individual components. In fact, his system would behave like ours only if the only constraints permitted are atomic constraints.

Whereas Brewka's system uses a recency-based heuristic to order belief constraints, our framework is more general by permitting arbitrary constraint prioritizations. We differ further from Brewka in that we factor out the use of priorities on beliefs entirely from the belief change process. Whereas Brewka's framework would only generate those maximal scenarios which respect the existing orderings on the beliefs, our framework would generate all maximal scenarios which satisfy the relevant belief constraints. Our framework would coincide with Brewka's, in this respect, if the constraint prioritization was based on recency.

Nayak *et al*[37], Boutilier [5] and Williams [51] address the question of generating a new selection function as result of a belief change step. However, they all use a recency-based heuristic for ranking revisions. More importantly, they do not address the problem of non-persistence of the effects of contractions.

## 4.5 *BR1*: Properties

In this section, we shall examine the extent to which System *BR1* satisfies the full preservation properties. We shall also test the system using the benchmark problems from the previous chapter.

Let us first examine compliance with the principle of irrelevance of syntax. We have seen in the previous section that the *BR1* system is independent of the syntactic form of the belief input. Thus, separately contracting two syntactically distinct but semantically equivalent sentences from the same PJ-default theory results in two equivalent PJ-default theories (in the sense that they have the same set of extensions). It is easy to see that the same holds for revision. In the AGM framework, as well as in the approaches described in the previous chapter, syntactically distinct but semantically equivalent representations of belief states are impossible since we always work with deductively closed theories. In System *BR1*, we manipulate finite sets of sentences, in a manner similar to Nebel's belief bases [40]. Syntactically distinct but semantically equivalent representations of belief states are thus possible. Hence, it is important to consider whether the outcome of a belief change step is contingent on the syntactic form of a belief state. It turns out that if we contract the same belief from two syntactically distinct but semantically equivalent PJ-default theories (in the sense that they have the same set of extensions), we get distinct outcomes. The following example illustrates this.

**Example 33** Consider the following two PJ-default theories:

$$(W_0, D_0) = (\{a, a \rightarrow b\}, \{\tfrac{:\top}{\top}\})$$
$$(W_0', D_0') = (\{a, b\}, \{\tfrac{:\top}{\top}\})$$

$(W_0, D_0)$ and $(W_0', D_0')$ are semantically equivalent. Both have exactly one extension, given by $Cn(\{a, b\})$.

Let us contract $b$ from $(W_0, D_0)$. The initial belief constraint set is $\{\{\{a\}, \{\neg a, b\}\}\}$. Since the belief constraint set consists of a single necessary belief constraint, the constraint prioritization relation is empty. The resulting belief state is given by:

$(W_1, D_1) = (\{\}, \{\frac{:a\wedge\neg b}{a}, \frac{:(a\rightarrow b)\wedge\neg b}{a\rightarrow b}, \frac{:\top\wedge b}{\top}\})$

Notice that $(W_1, D_1)$ has two extensions, given by $Cn(\{a\})$ and $Cn(\{a \rightarrow b\})$.

Let us now contract $b$ from $(W_0', D_0')$. The initial belief constraint set is $\{\{\{a\}, \{b\}\}\}$. Once again, since the belief constraint set consists of a single necessary belief constraint, the constraint prioritization relation is empty. The resulting belief state is given by:

$(W_1', D_1') = (\{a\}, \{\frac{:b\wedge\neg b}{b}, \frac{:\top\wedge\neg b}{\top}\})$

$(W_1', D_1')$ has a single extension, given by $Cn(\{a\})$.

Thus, after contracting the same belief from two semantically equivalent representations of a belief state, we obtain two semantically distinct belief states. □

We shall now consider the extent to which System *BR1* complies with the full preservation requirements. Here, the universe of beliefs, $\sum_+(S)$, consists of the union of the extensions of the PJ-default theory that denotes $S$. The universe of disbeliefs, $\sum_-(S)$, consists of the theory obtained by taking the conjunction of the elements of the set of disbelief constraints. Let us consider revision first. *Revision-1* requires that every belief that appears in some extension of the prior default theory appear in some extension of the revised default theory. This does not always hold in System *BR1*, as shown in the following example.

**Example 34** Let the initial belief state be given by:

$S_0 = (W_0, D_0) = (\{a\}, \{\frac{:\top}{\top}\})$

$(W_0, D_0)$ has a single extension, given by $Cn(\{a\})$. Let us revise with $\neg a$. The new constraint prioritization relation is given by:

$\{\{a\}\} \prec \{\{\neg a\}\}$

The resulting belief state is given by:

$S_1 = (W_1, D_1) = (\{\neg a\}, \{\frac{:a}{a}, \frac{:\top}{\top}\})$

$(W_1, D_1)$ has a single extension given by $Cn(\{\neg a\})$. Clearly, $a \in \sum_+(S_0)$ but $a \notin \sum_+(S_1)$.□

*Revision-2* is violated because beliefs previously suppressed by the existence of a disbelief constraint may reappear on account of this disbelief constraint being discarded, as shown in the following example.

**Example 35** Let the initial belief state be given by:

$$S_0 = (W_0, D_0) = (\{\}, \{\frac{:(a \wedge b) \wedge \neg a}{(a \wedge b)}, \frac{:\top \wedge \neg a}{\top}\})$$

$(W_0, D_0)$ has a single extension given by $Cn(\top)$. Let us revise $S_0$ with $a$. Let the new constraint prioritization relation be given by:

$$-\{\{\neg a\}\} \prec \{\{a\}\}$$

The resulting belief state is given by:

$$S_1 = (W_1, D_1) = (\{a\}, \{\frac{:(a \wedge b)}{(a \wedge b)}, \frac{:\top}{\top}\})$$

$(W_1, D_1)$ has a single extension given by $Cn(\{a, b\})$. $b \notin \sum_+(S_0)$, but $b \in \sum_+(S_1)$, even though $b$ is not a consequence of the new belief $a$ together with any element of $\sum_+(S_0)$.□

Example 35 also illustrates a case where *Revision-3* is violated. $\sum_-(S_0) = Cn(\{\neg a\})$, while $\sum_-(S_1) = Cn(\top)$. In general, disbelief constraints may be discarded when conflicting belief constraints of higher priority are added, hence *Revision-3* does not hold.

*Revision-4* does not hold in cases where the new necessary belief constraint is assigned a lower priority than an existing conflicting belief constraint.

**Example 36** Let us reverse the constrain prioritization in Example 34.

$$\{\{\neg a\}\} \prec \{\{a\}\}$$

In this case, the resulting belief state is given by:

$$S_1 = (W_1, D_1) = (\{a\}, \{\frac{:\neg a}{\neg a}, \frac{:\top}{\top}\})$$

$(W_1, D_1)$ has a single extension given by $Cn(\{a\})$. Clearly, the new belief $\neg a$ is not contained in $\sum_+(S_1)$.□

To consider compliance with *Revision-5* and *Revision-6*, let us view each belief state given by $(W, D)$ as a set of pairs $(e, d)$ where $e \in E((W, D))$ and $d$ is the theory obtained by taking the conjunction of the elements of $BC_{disbelief}$.

**Theorem 18** *System BR1 satisfies Revision-5 and Revision-6.*

**Proof:**

- *Revision-5:* Consider the THOERIST-translation of $(W, D)$. $d = Cn(C_{(W,D)}) = \sum_-(S)$. Then by definition of a maximal scenario, any superset is either unsatisfiable, or is not consistent with $d$. Hence proved.

- *Revision-6:* Since $d = \sum_-(S)$, the result holds.□

Consider, now, the full preservation conditions for contraction. It is easy to see why *Contraction-1* is violated. If the new disbelief constraint conflicts with a prior disbelief constraint, and the new constraint has higher priority, then the prior disbelief constraint is discarded. Hence, disbelief constraints are not preserved in general.

The introduction of a disbelief constraint can result in beliefs which appeared in some extension of the prior belief state being suppressed, as illustrated in Example 28, or in beliefs which were earlier suppressed appearing in some extension in the new belief state, as illustrated in Example 39. In either case, the new universe of beliefs is not identical to the prior universe of beliefs, hence *Contraction-3* is not satisfied in general.

As with revision, a contraction operation is not guaranteed to succeed. If a prior constraint, which conflicts with the new disbelief constraint, has a higher priority, then the new disbelief constraint is discarded. Hence *Contraction-4* is not satisfied in general.

**Theorem 19** *System BR1 satisfies Contraction-2, Contraction-5 and Contraction-6.*

**Proof:**

- *Contraction-2:* We know that $BC'_{disbelief} \subseteq BC_{disbelief} \cup \{x\}$ where $x$ is the new disbelief constraint. Since the universe of disbeliefs for any belief state is given by the conjunction of the elements of the set of disbelief constraints, the result follows.
- Contraction-5: The proof is identical to that for *Revision-5.*
- Contraction-6: The proof is identical to that for *Revision-6.*□

We shall now analyze the behaviour of the system using the benchmark problems we set up in the previous chapter.

**Example 37** Let us reformulate Example 1 from the previous chapter in System *BR1.* The initial belief state is given by:

$$S_0 = (W_0, D_0) = (\{\neg a\}, \{\frac{:\top}{\top}\})$$

Let us revise $S_0$ with $a$. Let the new constraint prioritization relation be given by:

$$\{\{\neg a\}\} \prec \{\{a\}\}$$

This prioritization corresponds to assigning a higher priority to the more recent belief input. The resulting belief state is given by:

$$S_1 = (W_1, D_1) = (\{a\}, \{\frac{:\neg a}{\neg a}, \frac{:\top}{\top}\})$$

We now contract $a$ from $S_1$. Let the new constraint prioritization relation assign, once again, a higher priority to the more recent belief input, as follows:

$$\{\{a\}\} \prec -\{\{\neg a\}\}$$

The resulting belief state is given by:

$$S_2 = (W_2, D_2) = (\{\}, \{\frac{:a \wedge \neg a}{a}, \frac{:\neg a \wedge \neg a}{\neg a}, \frac{:\top \wedge \neg a}{\top}\})$$

Notice that $(W_2, D_2)$ has a single extension given by $Cn(\{\neg a\})$. Thus, the original beliefs reappear after revising with $a$ and then undoing the revision (i.e. contracting $a$).□

**Example 38** Let us reformulate Example 2 from the previous chapter in System *BR1*. Note that we shall be taking belief base representations of the theories in Example 2. The syntactic form of the representation will therefore influence the outcome. Nevertheless, the example is useful in demonstrating the behaviour of the system.

Let the initial belief state be given by:

$$S_0 = (W_0, D_0) = (\{a, b\}, \{\tfrac{:\top}{\top}\})$$

Let the corresponding belief constraint set be given by $\{\{\{a\}\}, \{\{b\}\}\}$. Thus $a$ and $b$ constitute two distinct belief constraints.

We shall use a recency-based constraint prioritization throughout this example, i.e., the more recent input is assigned higher priority. Let us revise $S_0$ with $\neg b$. The new constraint prioritization is given by:

$$\{\{a\}\} \prec \{\{b\}\} \prec \{\{\neg b\}\}$$

The resulting belief state is given by:

$$S_1 = (W_1, D_1) = (\{a, \neg b\}, \{\tfrac{:b}{b}, \tfrac{:\top}{\top}\})$$

We now revise $S_1$ with $b$. The new constraint prioritization relation is given by:

$$\{\{a\}\} \prec \{\{\neg b\}\} \prec \{\{b\}\}$$

The resulting belief state is given by:

$$S_2 = (W_2, D_2) = (\{a, b\}, \{\tfrac{:\neg b}{\neg b}, \tfrac{:b}{b}, \tfrac{:\top}{\top}\})$$

Notice that the original beliefs $a$ and $b$ are contained in the sole extension of $(W_2, D_2)$ given by $Cn(\{a, b\})$. If the belief states had been represented in semantically equivalent but syntactically distinct forms, the outcome would have been different. Thus, if the initial $W_1$ was $\{a \wedge b\}$ instead of $\{a, b\}$, we would not have recovered the original beliefs after revising with $\neg b$ followed by revising with $b$.□

**Example 39** We shall go back to Example 3 in the previous chapter motivating the need to record contraction operations in the same way as revision operations. Let the initial belief state be given by :

The relation $\prec_1$ is empty.
$(W_1, D_1) = (\{b \rightarrow f\}, \{\frac{:\top}{\top}\})$
$BC_1 = \{\{\{\neg b, f\}\}\}$

The belief state obtained by contracting $f$ and assigning the highest priority to the disbelief constraint in $f$ is given by:

The relation $\prec_2$ is such that $\{\{\neg b, f\}\} \prec_2 -\{\{\neg f\}\}$
$(W_2, D_2) = (\{b \rightarrow f\}, \{\frac{:\top \wedge \neg f}{\top}\})$
$BC_2 = \{\{\{\neg b, f\}\}, -\{\{\neg f\}\}\}$

The belief state obtained by further revising with $b$ and assigning this belief constraint the highest priority is given by:

The relation $\prec_3$ is such that $\{\{\neg b, f\}\} \prec -\{\{\neg f\}\} \prec \{\{b\}\}$
$(W_3, D_3) = (\{b\}, \{\frac{:(b \rightarrow f) \wedge \neg f}{(b \rightarrow f)}, \frac{:\top \wedge \neg f}{\top}\})$
$BC_3 = \{\{\{\neg b, f\}\}, -\{\{\neg f\}\}, \{\{b\}\}\}$

$(W_3, D_3)$ has a single extension given by $Cn(\{b\})$. Notice that the contraction of $f$ persists since $f$ is in no extension of $(W_3, D_3)$.

If $\prec_3$ were such that $\{\{b\}\} \prec_3 \{\{\neg b, f\}\} \prec_3 -\{\{\neg f\}\}$, then the resulting belief state would be:

$$(W_3', D_3') = (\{b \rightarrow f\}, \{\frac{:b \wedge \neg f}{b}, \frac{:\top \wedge \neg f}{\top}\})$$

In this case the default theory would have a single extension $Cn(\{b \rightarrow f\}$.

If $\prec_3$ were such that $-\{\{\neg f\}\} \prec_3 \{\{b\}\} \prec_3 \{\{\neg b, f\}\}$, then the resulting belief state would be:

$$(W_3'', D_3'') = (\{b, b \rightarrow f\}, \{\frac{:\top}{\top}\})$$

The default theory would have a single extension $Cn(\{b, b \rightarrow f\})$.

Thus the three distinct candidate outcomes discussed in Example 3 are generated by taking different constraint prioritization relations. □

**Example 40** We shall reformulate Example 4 from the previous chapter in System *BR1*. Let the initial belief state be:

$$S_0 = (W_0, D_0) = (\{a, b\}, \{\frac{:\top}{\top}\})$$

Let the initial belief constraint set be $\{\{\{a\}\}, \{\{b\}\}\}$. Thus $a$ and $b$ constitute two distinct necessary belief constraints.

We first contract $b$ from $S_0$. Let the new constraint prioritization relation be:

$$\{\{a\}\} \prec \{\{b\}\} \prec -\{\{\neg b\}\}$$

The resulting belief state is given by:

$$S_1 = (W_1, D_1) = (\{a\}, \{\frac{:b \wedge \neg b}{b}, \frac{:\top \wedge \neg b}{\top}\})$$

We now revise $S_1$ with $c$. Let the new constraint prioritization be:

$$\{\{b\}\} \prec -\{\{\neg b\}\} \prec \{\{c\}\}$$

The resulting belief state is given by:

$$S_2 = (W_2, D_2) = (\{a, c\}, \{\frac{:b \wedge \neg b}{b}, \frac{:\top \wedge \neg b}{\top}\})$$

$(W_2, D_2)$ has a single extension $Cn(\{a, c\})$. The contracted belief $b$ does not reappear. In this case, this is a consequence of the representation of beliefs as finite sets of sentences rather than the explicit representation of the disbelief constraint for $b$.□

**Example 41** We reformulate Example 13 in System *BR1*. The initial belief state is empty:

$$S_0 = (W_0, D_0) = (\{\}, \{\frac{:\top}{\top}\})$$

After contracting $b$ from $S_0$, we obtain the following belief state:

$$S_1 = (W_1, D_1) = (\{\}, \{\frac{:\top \wedge \neg b}{\top}\})$$

We now contract $\neg b$ from $S_1$. If the new constraint prioritization relation is $-\{\{\neg b\}\} \prec -\{\{b\}\}$, then the resulting belief state is given by:

$$S_2 = (W_2, D_2) = (\{\}, \{\frac{:\top \wedge b}{\top}\})$$

If, on the other hand, the new constraint prioritization relation is $-\{\{b\}\} \prec -\{\{\neg b\}\}$, then the resulting belief state is given by:

$$S_2' = (W_2', D_2') = (\{\}, \{\frac{:\top \wedge \neg b}{\top}\})$$

Thus, in either case, one of the disbelief constraints is irretrievably lost. □

The previous example shows that System *BR1* suffers from the fundamental flaw in the constrained sets of theories belief representation scheme: multiple, mutually inconsistent sets of disbeliefs cannot be represented. However, unlike the operators defined in the constrained sets of theories approach, System *BR1* does not follow the strict AGM notion of success (note that it does not satisfy the *weak success* requirement either). Hence, it minimizes the discarding of beliefs, as seen in the reformulations of Examples 1 and 2. The ability to explicitly represent disbeliefs, as well as the ability to arbitrarily rank the belief constraints results in all candidate outcomes being considered in the reformulation of Example 3. System *BR1* also enjoys the benefits of representing beliefs as finite sets of sentences, as illustrated in the reformulation of Example 4.

## 4.6 *BR2*: Definition

System *BR2* was developed as a first step towards a finitely representable, implementable system that satisfied the full preservation requirements of the previous chapter. System *BR2* improves over System *BR1* by permitting multiple, mutually inconsistent sets of beliefs to be represented. This obviates the need for using a constraint prioritization relation in mapping from one belief state to another, although such a relation may still be used in the orthogonal task of theory preference in a given belief state.

System *BR2* corresponds to the sets of constrained theories approach described in the previous chapter, differing only in the fact that each belief or disbelief theory is represented as a finite set of sentences as opposed to a deductively closed theory. Recall that in the sets of constrained theories approach, a belief state is represented by a set of commitment states. A commitment state is given by a pair $(B, D)$ where $B$ is a theory denoting the current set of beliefs while $D$ is a theory denoting the current set of disbeliefs. Commitment states may be of two kinds:

- Maximal belief commitment states, where the belief theory is a maximal consistent subset of the universe of beliefs and the disbelief theory is a maximal subset of the universe of disbeliefs which is consistent with the belief theory.

- Maximal disbelief commitment states, where the disbelief theory is a maximal consistent subset of the universe of disbeliefs and the belief theory is a maximal subset of the universe of beliefs which is consistent with the disbelief theory.

In System *BR2*, a belief state is represented by a pair of PJ-default theories $((W_b, D_b), (W_d, D_d))$. The theory $(W_b, D_b)$ represents the set of maximal belief commitment states. The theory $(W_d, D_d)$ represents the set of maximal disbelief commitment states. Both $(W_b, D_b)$ and $(W_d, D_d)$ are *W-free*, i.e., the set of facts is necessarily empty. Let $\beta(\Delta)$ denote the set of consequents of all default rules in the set of default rules given by $\Delta$. Thus:

$$\beta(\Delta) = \{x \mid \frac{:x \wedge y}{x} \in \Delta\}$$

when $\Delta$ is a set of PJ-default rules. Let $\alpha(\Delta)$ denote the set consisting of the semi-normal part of the justifications of all default rules in the set of default rules given by $\Delta$. Thus:

$\alpha(\Delta) = \{y \mid \frac{:x \wedge y}{x} \in \Delta\}$

when $\Delta$ is a set of PJ-default rules. Every PJ-extension $(E_J, E_T)$ of $(W_b, D_b)$ denotes a maximal belief commitment state. $E_T$ denotes the disbelief theory. $E_J$ denotes the belief theory in an indirect way. We know that for every $E_J$, there exists a unique $y \in \alpha(D_b)$ such that $y \in E_J$. The theory $Cn(\{y\})$ represents the corresponding belief theory. $E_J$ does not represent a belief theory directly since, for every default rule $\frac{:x \wedge y}{x}$ fired in obtaining the PJ-extension containing $E_J$, both $x$ and $y$. Our intuition is that every distinct $y$ represents a maximal belief theory, while every distinct $x$ represents a candidate disbelief. Every PJ-extension $(E_J, E_T)$ of $(W_d, D_d)$ denotes a maximal disbelief commitment state. Here the roles of $E_J$ and $E_T$ are reversed. $E_J$ denotes the belief theory, while $E_T$ denotes the disbelief theory, once again, in an indirect way. For every distinct $E_J$, there exists a unique $y \in \alpha(D_d)$ such that $y \in E_J$. The theory $Cn(\{y\})$ represents the disbelief theory. Thus, the consequents of the default rules in $D_b$ denote potential beliefs, while the consequents of default rules in $D_d$ denote potential disbeliefs.

The PJ-default theories $(W_b, D_b)$ and $(W_d, D_d)$ will not, in general, be uniform, unlike the default theories used in System *BR1*. We shall see that each distinct semi-normal component of a justification of a default rule in $D_b$ represents a maximal disbelief theory, while each distinct semi-normal component of a justification of default rule in $D_d$ represents a maximal belief theory. We may partition a set of default rules into sets having an identical semi-normal component of the justification. We shall refer to this as the *justification-partitioning* (or *j-partitioning*). Formally, the *j-partitioning* of a set of PJ-default rules $\Delta$ is given by:

$P_\Delta = \{P_1, \ldots, P_n\}$

where each $P_i = \{\frac{:x_i \wedge y}{x_i} \mid \frac{:x_i \wedge y}{x_i} \in \Delta\}$

We may now define the operations of revision and contraction as follows:

*Revision:* Let $((W_b', D_b'), (W_d', D_d'))$ be the belief state obtained as a result of revising the belief state $((W_b, D_b), (W_d, D_d))$ with $x$. Let $P_{D_d} = \{P_1, \ldots, P_n\}$ be the j-partitioning of $D_d$. Then:

$$W'_d = \{\}$$
$$D'_d = \bigcup_i P'_i \text{ where}$$
$$P'_i = P_i \cup \{\frac{:x \wedge \alpha(P_i)}{x}\} \; W'_b = \{\}$$
$$D'_b = \{\frac{:y \wedge z}{y} \mid y \in \beta(D_b), z \in (\beta(D'_d) \downarrow \perp)\}$$

*Contraction:* Let $((W'_b, D'_b), (W'_d, D'_d))$ be the belief state obtained as a result of contracting $x$ from the belief state $((W_b, D_b), (W_d, D_d))$. Let $P_{D_b} = \{P_1, \ldots, P_n\}$ be the j-partitioning of $D_b$. Then:

$$W'_b = \{\}$$
$$D'_b = \bigcup_i P'_i \text{ where}$$
$$P'_i = P_i \cup \{\frac{:\neg x \wedge \alpha(P_i)}{\neg x}\} \; W'_d = \{\}$$
$$D'_d = \{\frac{:y \wedge z}{y} \mid y \in \beta(D_d), z \in (\beta(D'_b) \downarrow \perp)\}$$

In both revision and contraction, the set of facts in the resulting PJ-default theories are set to empty, since every PJ-default theory used in the representation of a belief state must necessarily be W-free. Revision begins with identifying the PJ-default theory $(W'_d, D'_d)$ representing the set of maximal disbelief commitment states. $W'_d$ is set to empty. A new default rule is added to each partition $P_i$ in the j-partitioning of $D_d$ with $\alpha(P_i)$ forming the semi-normal part of the justification of the default and the new belief forming the normal part of the justification and the consequent. In other words, for every distinct maximal consistent set of disbeliefs (the $\alpha(P_i)$ for each distinct $P_i$), a new default rule is created with the new belief as the consequent to obtain $D'_d$. Next, the default theory $(W'_b, D'_b)$ representing the set of maximal belief commitment states is identified. $W'_b$ is set to empty. For every distinct element of $(\beta(D'_d) \downarrow \perp)$ (i.e., for every maximal consistent set of beliefs), and for every default consequent in $D_d$ (i.e., for every distinct disbelief) a new default rule is created, to obtain $D'_d$. Contraction is the exact dual of revision. First, the new default theory $(W'_b, D'_b)$ representing the set of maximal belief commitment states is identified. A new default rule is added to each partition $P_i$ in the j-partitioning of $D_b$ with $\alpha(P_i)$

forming the semi-normal part of the justification of the default and the negation of the belief being contracted forming the normal part of the justification and the consequent. In other words, for every distinct maximal consistent set of beliefs (the $\alpha(P_i)$ for each distinct $P_i$), a new default rule is created with the negation of the belief being contracted as the consequent to obtain $D'_b$. Next, the default theory $(W'_d, D'_d)$ representing the set of maximal belief commitment states is identified. $W'_d$ is set to empty. For every distinct element of $(\beta(D'_b) \downarrow\perp)$ (i.e., for every maximal consistent set of disbeliefs), and for every default consequent in $D_d$ (i.e., for every distinct belief) a new default rule is created, to obtain $D'_d$.

No belief or disbelief is discarded in System *BR2*, hence no choice function, such as the one that uses the constraint prioritization relation in System *BR1*, is required. Belief change in System *BR2* involves only belief maintenance. Theory preference, i.e., the selection of a commitment state is a totally orthogonal task. All of the benefits of the irrelevance of choice, discussed in the previous chapter, accrue.

The computationally hard component of a belief change operation is the identification of maximal consistent subsets of a finite set of sentences. Thus, revision requires the computation of maximal consistent subsets of the new set of beliefs, while contraction requires the computation of maximal consistent subsets of the set of disbeliefs. We shall see in the next chapter that some straightforward, and practically viable special cases exist where this problem becomes tractable.

The system supports *lazy evaluation.* Part of the computation involved in identifying maximal commitment states from the representation of a belief state is performed during a belief change step (this corresponds to the computation of the maximal consistent subsets of the set of beliefs, or disbeliefs). The rest of the computation is deferred, and performed only when an agent needs to reason with, or act on, its beliefs. Actual commitment states, i.e., PJ-default extensions, need to be computed only when theory preference is required. As a consequence of representing the maximal belief, and maximal disbelief, commitment states in two distinct default theories, some of the computation may be deferred even at this stage. For instance, if an agent chooses to commit only to a maximal belief commitment state, then the PJ-extensions of only the theory $(W_b, D_b)$ from the pair of default theories $((W_b, D_b), (W_d, D_d))$ need to be computed.

## 4.7 *BR2*: Properties

In this section, we shall examine the behaviour of System *BR2* using the benchmark problems from the previous chapter. We shall formally establish the extent to which the system complies with the full preservation requirements.

**Example 42** Let us reformulate Example 1 from the previous chapter in System *BR2*. The initial belief state is given by:

$S0 = ((W0_b, D0_b), (W0_d, D0_d))$ where
$W0_b = \{\}$
$D0_b = \{\frac{:\top \wedge \neg a}{\top}\}$
$W0_d = \{\}$
$D0_d = \{\frac{:\neg a \wedge \top}{\neg a}\}$

Notice that $(W0_b, D0_b)$ has a single PJ-extension $(Cn(\{\neg a\}), Cn())$. This corresponds to a commitment state where the belief theory is given by $Cn(\{\neg a\})$ and the disbelief theory is given by $Cn()$. $(W0_d, D0_d)$ has a single PJ-extension $(Cn(), Cn(\{\neg a\}))$. This corresponds to an identical commitment state where the belief theory is given by $Cn(\{\neg a\})$ and the disbelief theory is given by $Cn()$.

After revising $S0$ with $a$, we obtain:

$S1 = ((W1_b, D1_b), (W1_d, D1_d))$ where
$W1_b = \{\}$
$D0_b = \{\frac{:\top \wedge \neg a}{\top}, \frac{:\top \wedge a}{\top}\}$
$W0_d = \{\}$
$D0_d = \{\frac{:\neg a \wedge \top}{\neg a}, \frac{:a \wedge \top}{a}\}$

Contracting $a$ from $S1$ we obtain:

$S2 = ((W2_b, D2_b), (W2_d, D2_d))$ where

$W2_b = \{\}$

$D2_b = \{\frac{:\top\wedge\neg a}{\top}, \frac{:\top\wedge a}{\top}, \frac{:\neg a\wedge\neg a}{\neg a}, \frac{:\neg a\wedge a}{\neg a}\}$

$W2_d = \{\}$

$D2_d = \{\frac{:\neg a\wedge\neg a}{\neg a}, \frac{:a\wedge\neg a}{a}\}$

$(W2_b, D2_b)$ has two PJ-extensions: $(Cn(\{\neg a\}), Cn(\{\neg a\})$ and $(Cn(\{a\}), Cn())$. This corresponds to two commitment states. The first has $Cn(\{\neg a\})$ as the belief theory and $Cn(\{\neg a\})$ as the disbelief theory. The second has $Cn(\{a\})$ as the belief theory and $Cn()$ as the disbelief theory. $(W2_d, D2_d)$ has a single PJ-extension given by $(Cn(\{\neg a\}), Cn(\{\neg a\})$ where the belief theory is given by $Cn(\{\neg a\})$ and the disbelief theory is given by $Cn(\{\neg a\})$ as well. This commitment state is identical to the first commitment state generated by $(W2_b, D2_b)$. Notice that the two commitment states in $S2$ are identical to the commitment states we obtain when we run this example using the sets of constrained theories approach. Note also that the initial belief $\neg a$ is contained in one of the belief theories in $S2$.□

**Example 43** Let us reformulate Example 2 from the previous chapter in System *BR2*. The initial belief state is given by:

$S0 = ((W0_b, D0_b), (W0_d, D0_d))$ where

$W0_b = \{\}$

$D0_b = \{\frac{:\top\wedge a\wedge b}{\top}\}$

$W0_d = \{\}$

$D0_d = \{\frac{:a\wedge\top}{a}, \frac{:b\wedge\top}{b}\}$

$(W0_b, D0_b)$ has a single PJ-extension $(Cn(\{a, b\}), Cn())$. This corresponds to a commitment state with a belief theory $Cn(\{a, b\})$ and a disbelief theory $Cn()$. $(W0_d, D0_d)$ has a single PJ-extension $(Cn(\{a, b\}), Cn(\{a, b\})$. This corresponds to a commitment state with a belief theory given by $Cn(\{a, b\})$ and a disbelief theory given by $Cn()$.

After revising $S0$ with $\neg b$, we obtain:

$S1 = ((W1_b, D1_b), (W1_d, D1_d))$ where

$W1_b = \{\}$

$D1_b = \{\frac{:\top \wedge a \wedge b}{\top}, \frac{:\top \wedge a \wedge \neg b}{\top}\}$

$W1_d = \{\}$

$D1_d = \{\frac{:a \wedge \top}{a}, \frac{:b \wedge \top}{b}, \frac{:\neg b \wedge \top}{\neg b}\}$

Further revising $S1$ with $b$, we obtain:

$S2 = ((W2_b, D2_b), (W2_d, D2_d))$ where

$W2_b = \{\}$

$D2_b = \{\frac{:\top \wedge a \wedge b}{\top}, \frac{:\top \wedge a \wedge \neg b}{\top}\}$

$W2_d = \{\}$

$D2_d = \{\frac{:a \wedge \top}{a}, \frac{:b \wedge \top}{b}, \frac{:\neg b \wedge \top}{\neg b}\}$

$(W2_b, D2_b)$ has two PJ-extensions: $(Cn(\{a, b\}), Cn())$ and $(Cn(\{a, \neg b\}), Cn())$. The first corresponds to a commitment state with a belief theory $Cn(\{a, b\})$ and a disbelief theory $Cn()$. The second corresponds to a commitment state with a belief theory $Cn(\{a, \neg b\})$ and a disbelief theory $Cn()$. $(W2_d, D2_d)$ has two PJ-extensions: $(Cn(\{a, b\}), Cn(\{a, b\}))$ and $(Cn(\{a, \neg b\}), Cn(\{a, \neg b\}))$. The first corresponds to a commitment state with a belief theory $Cn(\{a, b\})$ and a disbelief theory $Cn()$. The second corresponds to a commitment state with a belief theory $Cn(\{a, \neg b\})$ and a disbelief theory $Cn()$.

Notice that the original beliefs $a$ and $b$ appear in a belief theory in the final belief state. Notice also that the commitment states we obtain with System *BR2* are not identical to the those we obtain when we run the same example with the sets of constrained theories approach. The latter generates a superset of the set of commitment states we obtain here, as a consequence of working with deductively closed theories instead of finite sets of sentences. □

**Example 44** Let us reformulate Example 3 from the previous chapter in System *BR2*. The initial belief state is given by:

$S0 = ((W0_b, D0_b), (W0_d, D0_d))$ where
$W0_b = \{\}$
$D0_b = \{\frac{:\top \wedge (b \rightarrow f)}{\top}\}$
$W0_d = \{\}$
$D0_d = \{\frac{:(b \rightarrow f) \wedge \top}{(b \rightarrow f)}\}$

$(W0_b, D0_b)$ has single PJ-extension: $(Cn(\{b \rightarrow f\}), Cn())$. This corresponds to a commitment state with a belief theory $Cn(\{b \rightarrow f\})$ and a disbelief theory $Cn()$. $(W0_d, D0_d)$ has a single PJ-extension: $(Cn(\{b \rightarrow f\}), Cn(\{b \rightarrow f\}))$. This too corresponds to a commitment state with a belief theory $Cn(\{b \rightarrow f\})$ and a disbelief theory $Cn()$.

After contracting $f$ from $S0$ we obtain:

$S1 = ((W1_b, D1_b), (W1_d, D1_d))$ where
$W1_b = \{\}$
$D1_b = \{\frac{:\top \wedge (b \rightarrow f)}{\top}, \frac{:\neg f \wedge (b \rightarrow f)}{\neg f}\}$
$W1_d = \{\}$
$D1_d = \{\frac{:(b \rightarrow f) \wedge \neg f}{(b \rightarrow f)}\}$

Revising $S1$ with $b$, we obtain:

$S2 = ((W2_b, D2_b), (W2_d, D2_d))$ where
$W2_b = \{\}$
$D2_b = \{\frac{:\top \wedge (b \rightarrow f) \wedge b}{\top}, \frac{:\neg f \wedge (b \rightarrow f) \wedge b}{\neg f}\}$
$W2_d = \{\}$
$D2_d = \{\frac{:(b \rightarrow f) \wedge \neg f}{(b \rightarrow f)}, \frac{:b \wedge \neg f}{b}\}$

$(W2_b, D2_b)$ has a single PJ-extension: $(Cn(\{b \rightarrow f, b\}), Cn())$. This corresponds to a commitment state with a belief theory $Cn(\{b \rightarrow f, b\})$ and a disbelief theory $Cn()$. $(W2_d, D2_d)$ has two PJ-extensions: $(Cn(\{b \rightarrow f, \neg f\}), Cn(\{b \rightarrow f\}))$ and $(Cn(\{b, \neg f\}), Cn(\{b\}))$. The first corresponds to a commitment state with a belief theory $Cn(\{b \rightarrow f\})$ and disbelief theory $Cn(\{\neg f\})$. The second corresponds to a commitment state with a belief theory $Cn(\{b\})$ and a disbelief theory $Cn(\{\neg f\})$. Notice that each of the three candidate outcomes discussed in Example 3 appears as distinct belief theories in the final belief state. □

**Example 45** Let us reformulate Example 13 in System *BR2*. The initial belief state is given by:

$$S0 = ((W0_b, D0_b), (W0_d, D0_d)) \text{ where}$$
$$W0_b = \{\}$$
$$D0_b = \{\frac{:\top}{\top}\}$$
$$W0_d = \{\}$$
$$D0_d = \{\frac{:\top}{\top}\}$$

The initial belief state is thus empty.

After contracting $b$ from $S0$, we obtain:

$$S1 = ((W1_b, D1_b), (W1_d, D1_d)) \text{ where}$$
$$W1_b = \{\}$$
$$D1_b = \{\frac{:\top}{\top}, \frac{:\neg b \wedge \top}{\neg b}\}$$
$$W1_d = \{\}$$
$$D1_d = \{\frac{:\top \wedge \neg b}{\top}\}$$

Further contracting $\neg b$ from $S1$, we obtain:

$$S2 = ((W2_b, D2_b), (W2_d, D2_d)) \text{ where}$$
$$W2_b = \{\}$$
$$D2_b = \{\frac{:\top}{\top}, \frac{:\neg b \wedge \top}{\neg b}, \frac{:b \wedge \top}{b}\}$$
$$W2_d = \{\}$$
$$D2_d = \{\frac{:\top \wedge \neg b}{\top}, \frac{:\top \wedge b}{\top}\}$$

$(W2_b, D2_b)$ has two PJ-extensions: $(Cn(\{\neg b\}), Cn(\{\neg b\}))$ and $(Cn(\{b\}), Cn(\{b\}))$. The first corresponds to a commitment state with a belief theory $Cn()$ and a disbelief theory $Cn(\{\neg b\})$. The second corresponds to a commitment state with a belief theory $Cn()$ and a disbelief theory $Cn(\{b\})$. $(W2_d, D2_d)$ has two PJ-extensions: $(Cn(\{\neg b\}), Cn())$ and $(Cn(\{b\}), Cn())$. The first corresponds to a commitment state with a belief theory $Cn()$ and a disbelief theory $Cn(\{\neg b\})$. The second corresponds to a commitment state with a belief theory $Cn()$ and a disbelief theory $Cn(\{b\})$. Note that, unlike System *BR1*, the conflicting disbeliefs in $b$ and $\neg b$ appear in distinct disbelief theories in the final belief state. This stems from the facility in System *BR2* to retain conflicting disbeliefs, which is not possible in System *BR1*. □

Since no belief or disbelief is discarded in System *BR2*, the following observations can be made.

**Observation:** *Revision in System BR2 satisfies Revision-1 through Revision-4.*

**Observation:** *Contraction in System BR2 satisfies Contraction-1 through Contraction-4.*

Since System *BR2* operates on finite sets of sentences as opposed to deductively closed theories, *Revision-5* and *Revision-6*, as well as *Contraction-5* and *Contraction-6* are not satisfied. However, a weaker version of these conditions can be satisfied. The weakening involves taking a syntactic version of the universe of belief and disbeliefs. We shall refer to these syntax-based variants as the *potential belief base* and the *potential disbelief base.* The potential belief base $\sigma_+(S)$ of a belief state $S$ consists of all explicitly represented candidate beliefs in $S$ and is defined as follows:

$$\sigma_+(((W_b, D_b), (W_d, D_d))) = \{x \mid \frac{:x \wedge y}{x} \in D_d\}$$

Note that in the definition above, it is sufficient to refer to $D_d$ since every candidate belief will appear as the consequent of some default rule in $D_d$. The potential disbelief base $\sigma_-(S)$ of a belief state $S$ consists of all explicitly represented disbeliefs in $S$ and is defined as follows:

$$\sigma_-(((W_b, D_b), (W_d, D_d))) = \{x \mid \frac{:x \wedge y}{x} \in D_b\}$$

It is sufficient to refer to $D_b$ in the definition above since every candidate disbelief will appear as a consequent of some default rule in $D_b$.

**Observation:** *Revision in System BR2 satisfies Revision-5 and Revision-6 with the potential belief base and potential disbelief base substituted in place of the universe of beliefs and universe of disbeliefs respectively.*

**Observation:** *Contraction in System BR2 satisfies Contraction-5 and Contraction-6 with the potential belief base and potential disbelief base substituted in place of the universe of beliefs and universe of disbeliefs respectively.*

## 4.8 *BR2:* Related Work

System *BR1* improves on related approaches such as [6], [5], [37], [53] since none of these retain multiple sets of disbelief, and since most do not explicitly represent disbeliefs at all. In other respects, the observations made while comparing System *BR1* with these systems apply, since System *BR2* is a generalization of System *BR1*.

A few comments on the status of these systems with respect to the foundational versus coherentist taxonomy are in order. The line dividing these two epistemologies is thin, and it is often difficult to make a definitive classification of a given approach. Such is the case with Systems *BR1* and *BR2*. Both systems essentially take finite representations of beliefs as sets of sentences instead of deductively closed theories. This provides a foundational flavour, since the explicitly represented beliefs serve as the self-evident beliefs, while all derived beliefs rely on these for justification. On the other hand, the focus of these two systems is on maintaining coherence in the corpus of explicitly represented beliefs in a belief state, which provides a foundational flavour. In fact, the focus on maintaining justifications is minimal - it is limited to represented belief justifying derived ones.

# Chapter 5

# Implementation Strategies

Essential to the development of practical systems for belief change are strategies for efficient implementation. In the preceding chapters, we have presented a competence theory which provides a specification of belief change in the ideal case under the minimal change requirement. The competence theory accounts for practically motivated concerns, such as the need for defining belief change operators in the iterated case as well as the need to account for belief inputs which are uncertain or have low credibility. Providing, as it does, a specification of the ideal case, the competence theory is not computationally viable. We have presented the design of two systems, based on this competence theory, which are computationally viable. Both systems use the language of PJ-default logic to represent a belief state. We have identified some practical benefits of using such a belief representation language, such as the facility of using lazy evaluation approaches. Nevertheless, computing extensions of PJ-default theories is intractable in the general case. In this chapter, we shall outline strategies for efficiently implementing systems based on PJ-default logic, specially in situations where exponential time solutions are impractical. We shall present two classes of strategies. First, we shall define translations that permit the use of techniques from the area of partial constraint satisfaction in computing default extensions. Second, we shall define anytime procedures for computing partial extensions such that the quality of these partial extensions improve with time. The first set of techniques are useful both for the belief maintenance task, i.e., in obtaining a representation of the new belief state after a belief change step, as well as for the theory preference task, i.e., in identifying the applicable commitment state. Notice that in both Systems *BR1* and *BR2*, the belief maintenance task involves computing maximal consistent subsets of sets of sentences, which reduces to the problem of computing default extensions. As well, the theory preference task necessarily requires that commitment states, i.e., default extensions, be computed. The second set of techniques is useful only for the theory preference task, since partial solutions can affect the correctness of subsequent belief change steps in the case of the belief maintenance task. Both classes of techniques presented in this chapter are independently useful in a variety of information processing applications.

Default reasoning and the constraint satisfaction problem (CSP) formalization have independently emerged as major problem solving paradigms in AI and related areas. Very little has been done, however, to relate these two seemingly unrelated areas of inquiry. Recently, researchers have begun looking at situations, within the CSP paradigm, in which the problem is over-constrained (and thus admits no complete solution) or where resource-bounds do not permit the identification of complete solutions. Such problems are often termed as partial constraint satisfaction (PCSP) problems. Our first step in this chapter will be to identify common structure between the constraint satisfaction problem and the problem of default inference by reformulating propositional default inference as a problem of partial constraint satisfaction. Establishing such a connection is useful in several ways:

- A large corpus of results from the PCSP area can be directly applied for designing efficient algorithms for default inference.
- Tractable cases for PCSP suggest tractable classes of default theories
- Most PCSP techniques are suitable for resource-bounded computation and can form the basis for algorithms for resource-bounded default inference

Since default inference is inherently computationally hard [30] and practical applications, especially time-bounded ones, may require that some notion of approximate inference be used. Any approximation algorithm must provide useful partial results and the trade-offs involved must be clearly identified. Approximate default inference has received scant attention in the literature (the notable exception being the work of Cadoli and Schaerf [8] in which they improve on the complexity of reasoning with Reiter's default logic by using consequence relations that are sound and incomplete in one case and complete but unsound in the other).

Real-time algorithms are usually designed to satisfy a variety of application-specific requirements: some are required to provide partial, but useful results whenever they are stopped while others have the additional requirement that their partial results improve with time. *Anytime algorithms* are a useful conceptualization of processes that may be prematurely terminated whenever necessary to return useful partial answers, with the quality of the answers improving in a well-defined manner with time. Dean and Boddy [12] define an anytime algorithm to be one which:

- Lends itself to preemptive scheduling techniques.
- Can be terminated at any time and will give some meaningful answer.
- Returns answers that improve in some well-behaved manner as a function of time.

In the rest of this chapter, we shall informally refer to the process of an anytime algorithm progressively computing solutions of measurably improving quality as the *anytime progression.*

Here, we shall develop a repertoire of meaningful partial solutions for default inference problems and use these as the basis for specifying general classes of anytime default We shall then exploit the connection between PCS problems and default inference to identify a large space of possible algorithms for default inference that may be based on partial constraint satisfaction techniques, which are inherently anytime in nature. In effect, these algorithms will permit us to compute partial commitment states with a well-specified measure of the distance of the partial commitment state from the exact commitment state, together with the guarantee that this distance will get monotonically shorter as more time is spent identifying a commitment state. Notice that this set of techniques is useful only in the theory preference component of belief change. Using partial computation in the belief maintenance component can affect the correctness of subsequent belief state representations.

In the rest of this chapter, we shall refer to the $E_T$ component of a PJ-extension as an *extension.*

Part of the material on the connection between PCS problems and default inference has appeared in [26].

## 5.1 Partial Constraint Satisfaction

Partial constraint satisfaction techniques [18] (or PCS techniques) were developed for handling overconstrained constraint satisfaction problems. In other words, when a constraint satisfaction problem admits no solution that satisfies all the specified constraints, a partial constraint satisfaction technique will enable us to identify the "best" partial solution, where the notion of "best" can be defined using a variety of metrics. For example, one notion of the "best" partial solution could be the solution that satisfies the maximal number of constraints. Partial constraint satisfaction with this metric is often termed as *maximal constraint satisfaction.* Given the difficulty of obtaining a priori guarantees on the existence of solutions that satisfy all constraints, it is clear that PCS techniques have a broader applicability than classical CSP techniques. PCS techniques are also suitable for solving problems in resource-bounded situations, such as when the time available for computing a solution is bounded. PCS techniques can help us identify solutions that are "good enough" or "close enough" to the complete solution in the available time.

Formally, a partial constraint satisfaction problem specification consists of a finite set of variables $X = \{X_1, \ldots, X_n\}$, each associated with a domain of discrete values, $D_1, \ldots, D_n$, and a set of constraints $C = \{C_1, \ldots, C_m\}$. Each constraint is a relation defined on some subset of the set of variables. Formally, a constraint $C_i$ consists of the *constraint-subset* $S_i = \{X_{i_1}, \ldots, X_{i_{j(i)}}\}$, where $S_j \subseteq X$, denoting the subset of the variables on which $C_i$ is defined and the *relation* $rel_i$ defined on $S_i$ such that $rel_i \subseteq D_{i_1} \times \ldots \times D_{i_{j(i)}}$. We shall define a solution of a maximal constraint satisfaction problem ( or a *maximal solution*) to be an assignment of legal values (i.e. values from the respective domains) to each $X_i \in X$ such that the set of satisfied constraints $C' \subseteq C$ is such that there exists no $C''$ where $C' \subset C'' \subseteq C$ which is also satisfied by an assignment of legal values to all variables.

In this paper, we shall consider a variant of the standard PCSP formulation, which we shall refer to as *prioritized-PCSP*. In addition to the standard PCSP specification described above, a prioritized-PCSP specification contains a partial order $\succeq$ defined on the set $C$ of constraints. We say that a constraint $c$ is *preferred* to a constraint $c'$ iff $c \succeq c'$. If for some $c \in C$ there exists no $c' \in C$ such that $c' \succeq c$, then $c$ is referred to as an *essential constraint.* All elements of $C$ which are not essential constraints are referred to as *potential constraints.* Every solution to a prioritized-PCS problem must satisfy all essential constraints, although potential constraints may be violated. A solution $s$ to a prioritized-PCSP is said to be *preferred* over another solution $s'$ iff there exists a constraint c satisfied by $s$ and a constraint $c'$ satisfied by $s'$ such that c is preferred to $c'$ and there exists no $c''$ satisfied by $s'$ and $c'''$ satisfied by $s$ such that $c''$ is preferred to $c'''$. A solution $s$ is said to be *dominant* iff there exists no other solution $s'$ such that $s'$ is preferred over $s$.

## 5.2 Mapping default theories to PCS problems

In this section, we shall summarize our previous work [26] on translations from THEORIST systems and PJ-default theories to prioritized-PCSP specifications. We shall then present results relating dominant maximal solutions of these translations to the corresponding THEORIST or PJ-default extensions. The simplest case is the translation of a THEORIST system.

**Definition 11** [26] *Let $PCSP_{(F,H,C)}$ denote the translation of a THEORIST system $(F, H, C)$ to a prioritized-PCSP specification. $PCSP_{(F,H,C)}$ is a 4-tuple $< X_{(F,H,C)}, D_{(F,H,C)}, C_{(F,H,}$ where $X_{(F,H,C)}$ is the set of variables, $D_{(F,H,C)}$ is the set of domains for each of the variables, $C_{(F,H,C)}$ is the set of constraints and $\succeq_{(F,H,C)}$ is the partial order defined on $C_{(F,H,C)}$.*

- $X_{(F,H,C)} = Herbrand\text{-}Base(F \cup H \cup C)$.
- $D_{(F,H,C)} = \{D_1, \ldots, D_n\}$ *where each* $D_i = \{True, False\}$.
- $C_{(F,H,C)} = \{\Omega(x) \mid x \in F\} \cup \{\Omega(x) \mid x \in H\} \cup \{\Omega(x) \mid x \in C\}$,
  *where $\Omega(x)$, for some propositional sentence $x$ is a relation on the propositional letters in $x$ such that each tuple in the relation represents a truth assignment to these propositional letters which satisfies $x$.*
- $\forall c_1, c_2 \in C_{(F,H,C)}$, $c_1 \succeq c_2$ *iff:*
  1. $c_1 = \Omega(x)$ *s.t.* $x \in F$, $c_2 = \Omega(y)$ *s.t.* $y \in H$, *OR*
  2. $c_1 = \Omega(x)$ *s.t.* $x \in C$, $c_2 = \Omega(y)$ *s.t.* $y \in H$.

The translation involves treating each propositional letter in the Herbrand base of the THEORIST system as a variable for the prioritized-PCSP. Each constraint $\Omega(x)$ can be viewed as the relation corresponding to the satisfying truth assignments in the truth-table for the propositional sentence $x$. The partial order assigns a higher priority to constraints obtained from elements of $F$ and $C$ over those obtained from elements of $H$. Thus the former constitute the essential constraints, while the latter correspond to the potential constraints. Every solution will satisfy constraints obtained from $F$ and $C$, but not every constraint obtained from $H$.

**Theorem 20** [26] *Let $c_s$ denote the subset of $C_{(F,H,C)}$ that is satisfied by a dominant maximal solution of $PCSP_{(F,H,C)}$. Let $h_{c_s} = \{x \mid x \in H, \Omega(x) \in c_s\}$. Then $F \cup h_{c_s}$ is a maximal scenario for the THEORIST system $(F, H, C)$.*

**Proof outline:** Assume there exists some $h_{c_s}$ such that $F \cup h_{c_s}$ is not a maximal scenario. Since $h_{c_s} \subseteq H$, two cases are possible:

1. There exists some $h$ where $h_{c_s} \subset h$ such that $F \cup h$ is a maximal scenario. Consider some $x \in h - h_{c_s}$. Since $c_s$ is satisfied by a maximal solution and since $x \cup h_{c_s}$ is satisfiable, it must be true that $x \in c_s$. This contradicts our assumption.

2. $F \cup h_{c_s} \models \perp$. This is impossible since the constraints corresponding to $h_{c_s}$ are satisfied in the dominant maximal solution.

A similar translation can be defined for PJ-default theories.

**Definition 12** [26] *Let $PCSP_{(W,D)}$ denote the translation of a PJ-default theory $(W, D)$ to a prioritized-PCSP specification. $PCSP_{(W,D)}$ is a 4-tuple $< X_{(W,D)}, D_{(W,D)}, C_{(W,D)}, \succeq_{(W,$ where $X_{(W,D)}$ is the set of variables, $D_{(W,D)}$ is the set of domains for each of the variables, $C_{(W,D)}$ is the set of constraints and $\succeq_{(W,D)}$ is a partial order defined on $C_{(W,D)}$.*

- $X_{(W,D)}$ =*Herbrand-Base*$(W \cup \{\alpha \mid \frac{:\alpha\wedge\beta}{\alpha} \in D\} \cup \{\beta \mid \frac{:\alpha\wedge\beta}{\alpha} \in D\})$.
- $D_{(W,D)} = \{D_1, \ldots, D_n\}$ *where each* $D_i = \{True, False\}$.
- $C_{(W,D)} = \{\Omega(x) \mid x \in W\} \cup \{\Omega(\alpha \wedge \beta) \mid \frac{:\alpha\wedge\beta}{\alpha} \in D\}$,
  *where $\Omega(x)$ is as defined in Definition 2.*
- $\forall c_1, c_2 \in C_{(W,D)}$, $c_1 \succeq_{(W,D)} c_2$ *iff* $c_1 = \Omega(x)$ *s.t.* $x \in W$ *and* $c_2 = \Omega(\alpha \wedge \beta)$ *s.t.* $\frac{:\alpha\wedge\beta}{\alpha} \in D$.

**Theorem 21** [26] *Let $c_s$ denote the subset of $C_{(W,D)}$ that is satisfied by a dominant maximal solution of $PCSP_{(W,D)}$. Let $h_{c_s} = \{\alpha \mid \frac{:\alpha\wedge\beta}{\alpha} \in D, \Omega(\alpha \wedge \beta) \in c_s\}$. Then $Th(W \cup h_{c_s})$ is an extension for the PJ-default theory $(W, D)$.*

**Proof:** Follows from Theorem 14 and Theorem 20.

The prioritized-PCSP formulation permits us to extend these results for systems with default priorities such as Brewka's *preferred subtheories approach* [7]. Let a *prioritized THEORIST system* $(F, H, C, \geq)$ be a THEORIST system $(F, H, C)$ with a partial order $\geq$ defined on $H$ such that for $h, h' \in H$, $h$ is said to be preferred over $h'$ if $h \geq h'$. A scenario $S$ is preferred over another scenario $S'$ if $S$ contains a hypothesis that is preferred over some hypothesis contained in $S'$ and $S'$ does not contain any hypothesis that is preferred over some hypothesis contained in $S$. A dominant scenario is one for which there exists no other scenario that is preferred over it. Consider a prioritized-PCSP translation of this system such that the ordering $\succeq_{(F,H,C)}$ is extended such that a constraint obtained from $h$ is preferred over a constraint obtained from $h'$ if $h$ is preferred over $h'$. We do not present the full translation here for brevity, but it is easy to see that the following result holds.

**Observation:** Let $c_s$ denote the subset of $C_{(F,H,C,\geq)}$ that is satisfied by a dominant maximal solution of $PCSP_{(F,H,C,\geq)}$. Let $h_{c_s} = \{x \mid x \in H, \Omega(x) \in c_s\}$. Then $F \cup h_{c_s}$ is a dominant maximal scenario for the THEORIST system $(F, H, C)$.

Given these translations, the basic branch-and-bound algorithm for maximal constraint satisfaction presented by Freuder and Wallace in [18] can be applied for computing default extensions with a few minor modifications. The Freuder-Wallace branch-and-bound algorithm involves a backtrack search through the space of possible solutions. Three working parameters are maintained: the current best solution (i.e. the variable instantiation that violates the least number of constraints), the current *sufficient bound* (if a solution violating this or a fewer number of constraints is obtained, it is considered to be good enough and search stops) and the current *necessary bound* (this gives the number of constraints violated by the current best solution - if a branch of the search tree turns out to violate a higher number of constraints, then it is eliminated from consideration). An a priori value for the sufficient bound may be known (for example, if it is known that no solution violating less than 2 constraints is possible, the sufficient bound can be set to 2); otherwise, the sufficient bound is initially set to 0. An a priori value for the necessary bound may also be known (for instance, we may not be willing to settle for a solution violating more than $n$ constraints); otherwise its value is initially set to infinity. Unlike backtrack search for standard CSP, the Freuder-Wallace branch-and-bound algorithm does not treat a constraint violation as a failure of a search path; instead, it continues expanding a search path until the necessary bound is violated and returns the current best solution at the end of search. The modification that we require involves redefining the notion of acceptable constraint violations based on the partial order defined on the set of constraints. Whereas the Freuder-Wallace algorithm permits violation of any constraint, our modification permits the violation of only the *potential constraints*. A violation of an *essential constraint* results in the failure of the current search path, as in standard CSP, whereas the violation of a potential constraint is tolerated as long as the necessary bound is not exceeded. Default extensions are obtained by flagging each satisfied potential constraint while computing a maximal solution.

**Example 46** *Consider a PJ-default theory $(W, D)$ where:*

- $W = \{a\}$
- $D = \{d_1, d_2\}$ *where* $d_1 = \frac{:b \wedge \neg c}{b}$ *and* $d_2 = \frac{:c \wedge \neg b}{c}$.

*The corresponding translation $PCSP_{(W,D)}$ is as follows:*

- $X_{(W,D)} = \{a, b, c\}$.
- $D_x = \{T, F\}$ *for every* $x \in X_{(W,D)}$.
- $C_{(W,D)} = \{c_1, c_2, c_3\}$ *where:*
  - $c_1$ *is a relation defined on* $\{a\}$ *and consists of a single tuple* $\{T\}$.
  - $c_2$ *is a relation defined on* $\{b, c\}$ *and consists of a single tuple* $\{T, F\}$.
  - $c_3$ *is a relation defined on* $\{b, c\}$ *and consists of a single tuple* $\{F, T\}$.
- $c_1 \succeq c_2$ *and* $c_1 \succeq c_3$.

*$(W, D)$ has two extensions $Th(a, b)$ and $Th(a, c)$. $PCSP_{(W,D)}$ has two dominant maximal solutions: $\{a = T, b = T, c = F\}$ and $\{a = T, b = F, c = T\}$. In the first solution, constraint $c_1$ corresponding to default $d_1$ is satisfied while $c_2$ corresponding to $d_2$ is not. In the second solution, the reverse is true. Thus it is easy to see why Theorem 3 holds.*

Freuder and Wallace [18] have identified a maximal constraint satisfaction algorithm that is linear in the number of variables and quadratic in the maximum domain size for PCSP's with tree-structured constraint graphs, extending a previous similar result for tree-structured CSP's. Thus for default theories (either PJ-default theories or THEORIST systems) whose prioritized-PCSP translations have tree-structured constraint graphs, an extension can be identified in time linear in the number of propositional letters in the Herbrand base of the theory, since the domain size is constant for every variable and equals 2.

**Theorem 22** [26] *A default extension can be computed in $O(n)$ time, for default theories (either PJ-default theories or THEORIST systems) with prioritized-PCSP translations whose constraint graphs are tree-structured, where $n$ is the cardinality of the Herbrand base of the default theory.*

**Proof:** The proof follows directly from the observation that the Freuder and Wallace algorithm computes a default extension in the case the constraint graph corresponds to the PCSP translation of a default theory.□

## 5.3 A Framework for Anytime Default Inference

### 5.3.1 Partial Solutions

The essence of an anytime algorithm is to exploit some measure of progress towards a goal, in order to estimate the quality of a current solution at any time during the computation. Within the framework of nonmonotonic reasoning, there are several possible formulations of problem and solution. For example, Lin and Goebel [33] propose six different methods of using hypothetical reasoning to produce six different kinds of prediction, and show transformations to circumscription based on MILO resolution [43]. They distinguish, for example, between a weak prediction that is a consequence of only one extension of a default theory, versus a formula that is a consequence of all extensions of that same default theory.

Similarly, we here propose a simpler but useful set of nonmonotonic reasoning solutions, which provide us with a range of problem-solving behaviours over which anytime algorithms can be defined. We shall consider *solutions* to *default inference problems* in the following sense.

**Definition 13** *A* solution *for a default inference problem, given a default theory* $(W, D)$*, is an answer to any of the following queries:*

- Coherence: *Compute an extension of* $(W, D)$.
- Set-membership: *Determine if a given formula is an element of any extension of* $(W, D)$.
- Set-entailment: *Determine if a given formula is an element of all extensions of* $(W, D)$.

Note that *set-membership* is a kind of reckless version of the more conservative *set-entailment.* All three notions of solution distinguish different potential problem-solving behaviours. One could use *coherence* as the specification for the computation of belief states (cf. [23] [25]). As with Lin and Goebel [33], *set-membership* and *set-entailment* can be used as the basis for many different kinds of problem solving, including diagnosis and prediction.

In the context of propositional default inference, there are at least two immediately obvious ways of defining partial solutions: those which consider only a subset of the set of propositional letters used in the given default theory, and those which look at a subset of the default theory itself. We shall call these *$\alpha$-partial solutions* and *$\beta$-partial solutions* respectively. A partial solution may potentially fail to take into consideration the entire set of facts $W$ in a default theory $(W, D)$. Informally, we shall call a partial solution *W-preserving* only if it takes into account all the constraints imposed by $W$ in a default theory $(W, D)$. We shall be primarily interested in W-preserving solutions, but shall point out situations where useful classes of solutions can be W-preserving only in a limited sense.

To formalize $\alpha$-partial solutions, we need to define a precise notion of what it means to *restrict* a default theory based on a set of propositional letters. In the following, $\mathcal{P}$ stands for the set of propositional letters in the language $\mathcal{L}$. $\lambda((W,D))$ refers to the set of propositional letters appearing in the default theory $(W,D)$. $D_{\mathcal{P}}$ stands for the set of possible default theories $(W,D)$ such that $\lambda((W,D)) \subseteq \mathcal{P}$. We shall assume that every element of $W$ as well the justifications and consequents of every default rule in $D$ is written in conjunctive normal form.

**Definition 14** *A* restriction *of a default theory is a function* $\mathcal{R}$: $D_{\mathcal{P}} \times 2^{\mathcal{P}} \rightarrow D_{\mathcal{P}}$ *such that* $\lambda(R((W,D),S)) \subseteq S$ *where* $(W,D) \in D_{\mathcal{P}}$ *and* $S \in 2^{\mathcal{P}}$, *where* $\mathcal{R}$ *can take one of the two following forms:*

- *If* $\mathcal{R}$ *is a* strong restriction, *then for every disjunction* $\phi$ *s.t.* $\phi \in W$ *or* $\phi \in \beta$ *where* $\frac{:\beta}{\gamma} \in D$ *or* $\phi \in \gamma$ *where* $\frac{:\beta}{\gamma} \in D$, *if* $\phi = \alpha_1 \vee \alpha_2 \vee \ldots \vee \alpha_n$ *and if there exists any* $\alpha_i$ *s.t.* $\alpha_i \notin S$, *if* $\alpha_i$ *is a positive literal or* $\neg\alpha_i \notin S$, *if* $\alpha_i$ *is a negative literal, then* $\phi$ *is replaced by* $\top$ *in* $\mathcal{R}((W,D),S)$.
- *If* $\mathcal{R}$ *is a* weak restriction, *then for every disjunction* $\phi$ *s.t.* $\phi \in W$ *or* $\phi \in \beta$ *where* $\frac{:\beta}{\gamma} \in D$ *or* $\phi \in \gamma$ *where* $\frac{:\beta}{\gamma} \in D$, *if* $\phi = \alpha_1 \vee \alpha_2 \vee \ldots \vee \alpha_n$ *and for every* $\alpha_i$ *s.t.* $\alpha_i \notin S$, *if* $\alpha_i$ *is a positive literal or* $\neg\alpha_i \notin S$, *if* $\alpha_i$ *is a negative literal,* $\phi$ *is replaced by* $\alpha_1 \vee \alpha_2 \vee \ldots \alpha_{i-1} \vee \alpha_{i+1} \vee \ldots \vee \alpha_n$ *in* $\mathcal{R}((W,D),S)$.

Weak and strong restrictions reflect two distinct intuitions on what it means to consider a part of a theory that relates to a subset of the set of propositional letters appearing in a theory. Strong restrictions reflect the intuition that if a propositional letter, whose truth status we are indifferent to, appears in a disjunction, then the disjunction evaluates to true by virtue of our indifference sanctioning the assumption that the propositional letter (or its negation, if it appears in the disjunction in negated form) evaluates to true. Weak restrictions reflect the intuition that if a propositional letter appears in a disjunction (in positive or negated form), then we are interested in the portion of the disjunction that does not involve this propositional letter. It appears that there are no first principles argument for preferring one form of restriction over the other. An actual choice will probably be driven by application-specific considerations.

**Definition 15** *An* $\alpha$-partial solution *with respect to a set* $S \subseteq \mathcal{P}$ *for a default inference problem given a default theory* $(W, D)$ *is a* solution *for a default inference problem given a default theory* $\mathcal{R}((W, D), S)$.

An $\alpha$-partial solution is thus a solution which considers only those portions of the default theory which involve some subset of the set of propositional letters which appear in the original default theory. An *$\alpha$-partial default inference procedure*, when prematurely terminated, would return an $\alpha$-partial solution with respect to the subset of the propositional letters it has been able to consider this far. Such a procedure would return better solutions if it is stopped later, since it would be able to consider a larger subset of the set of propositional letters. A priority relation defined on $\lambda((W, D))$ could define the anytime progression, with higher priority propositional letters being considered before lower priority ones. $\alpha$-partial procedures are *W-preserving* in a weak sense, since they only consider $W$ restricted to the set of propositional letters of interest.

**Definition 16** *A* $\beta$-partial solution *to a default inference problem, given a default theory* $(W, D)$ *is a* solution *for a default inference problem given a default theory* $(W', D')$ *where either* $W' \subseteq W$ *or* $D' \subseteq D$ *or both.*

Thus, a $\beta$-partial solution is one which looks at some subset of the default theory, while considering the entire set of propositional letters. Usually, only W-preserving solutions are of interest, so only subsets of $D$ are considered. A *$\beta$-partial default inference procedure*, when prematurely terminated, would return a solution which respects some subset of $D$. Such a procedure would return better solutions if it is stopped later, since it would be able to consider a larger subset of the set of the set of defaults. A priority relation defined on $(W, D)$ could define the anytime progression, with higher priority defaults being considered before lower priority ones.

A third variety of partial solution is conceivable: those which consider a subset of the set of propositional letters as well as a subset of the set of defaults. We shall refer to these as $\alpha\beta$-partial solutions.

**Definition 17** *An $\alpha\beta$-partial solution with respect to a set $S \subseteq \mathcal{P}$ and a set of defaults $D' \subseteq D$ for a default inference problem given a default theory $(W, D)$ is a* solution *for a default inference problem given a default theory $\mathcal{R}((W, D'), S)$.*

### 5.3.2 Anytime procedures

In this section, we present a set of generalized procedures that return $\alpha$ and $\beta$-partial solutions to coherence, set-membership and set-entailment queries. These procedures lay the foundations on which actual strategy for partial solution computation should be based. The formal properties of these procedures, stated at the end of this section, explicate the precise nature of rationality guarantees that each of the two classes of partial solutions provide. The performance of these procedures can be optimized in a variety of general as well as implementation-specific ways. However, we do not go into such details here, since our purpose is merely to present the general nature of partial solution computation strategies.

We assume that a procedure called COMPUTE-EXTENSION exists, which takes a default theory and a sequence of default rules and returns exactly one extension, obtained by attempting to fire each default rule in the sequence provided. The idea is simply to refer to some machinery for generating extensions, with a guarantee that all extensions will be eventually generated (this is achieved by providing different sequences of default rules in the input). A variety of techniques for doing this have been described in the literature on default reasoning, and we shall therefore not describe this procedure in any further detail. The input sequence of propositional letters in the ALPHA class of algorithms determines the order in which these letters are considered. The input sequence of default rules in the BETA class of algorithms similarly determines the order in which progressively larger subsets of default rules are considered. Notice that the procedures defined in this section do not require any commitment to weak or strong restriction functions for default theories. That decision can be guided by application-specific considerations.

**Algorithm 1** *ALPHA-COHERENCE*

**Input:** *A default theory (W,D), a logical variable INTERRUPT and a sequence* $< l_1, l_2, \ldots, l_n >$ *containing every element of* $\lambda((W, D))$.

**Output:** *A set of formulas EXTENSION and a set of propositional letters* $S \subseteq \lambda((W, D))$.

> $S = \{l_1\};\ i = 1$
> *do while NOT(INTERRUPT) and* $S \subset \lambda((W, D))$
>   $S = S \cup \{l_{i+1}\};\ i = i + 1.$
>   *EXTENSION=COMPUTE-EXTENSION*$(\mathcal{R}((W, D), S), < d_1, d_2, \ldots, d_m >)$
>   *where* $< d_1, d_2, \ldots, d_m >)$ *is any arbitrary permutation of the elements of D*
> *return* $< EXTENSION, S >$
> *stop*

**Algorithm 2** *ALPHA-SET-MEMBERSHIP*

**Input:** *A default theory (W,D), a logical variable INTERRUPT, a formula F and a sequence* $< l_1, l_2, \ldots, l_n >$ *containing every element of* $\lambda((W, D))$.

**Output:** *A logical variable IN-EXTENSION and a set of propositional letters* $S \subseteq \lambda((W, D))$.

> $S = \{l_1\};\ i = 1$
> *do while NOT(INTERRUPT) and* $S \subset \lambda((W, D))$
>   $S = S \cup \{l_{i+1}\};i = i + 1$
>   *for every permutation* $< d_1, d_2, \ldots, d_m >)$ *of the elements of D*
>     *EXTENSION=COMPUTE-EXTENSION*$(\mathcal{R}((W, D), S), < d_1, d_2, \ldots, d_m >)$
>     *if* $\{\neg F\} \cup$*EXTENSION is unsatisfiable then*
>     *return* $< IN - EXTENSION{=}TRUE, S >$
>     *stop*

*return* $< IN - EXTENSION{=}FALSE, S >$

*stop*

**Algorithm 3** *ALPHA-SET-ENTAILMENT* **Input:** *A default theory (W,D), a logical variable INTERRUPT, a formula F and a sequence* $< l_1, l_2, \ldots, l_n >$ *containing every element of* $\lambda((W, D))$.

**Output:** *A logical variable IN-ALL-EXTENSIONS and a set of propositional letters* $S \subseteq \lambda((W, D))$.

$S = \{l_1\}$; $i = 1$

*do while NOT(INTERRUPT) and* $S \subset \lambda((W, D))$

$S = S \cup \{l_{i+1}\}$;$i = i + 1$

*IN-ALL-EXTENSIONS=TRUE*

*for every permutation* $< d_1, d_2, \ldots, d_m >)$ *of the elements of D*

*EXTENSION=COMPUTE-EXTENSION*$(\mathcal{R}((W, D), S), < d_1, d_2, \ldots, d_m >)$

*if* $\{\neg F\}\cup$*EXTENSION is satisfiable then*

*IN-ALL-EXTENSIONS=FALSE*

*return* $< IN - ALL - EXTENSIONS, S >$

*stop*

**Algorithm 4** *BETA-COHERENCE*

**Input:** *A default theory (W,D), a logical variable INTERRUPT and a sequence* $< d_1, d_2, \ldots, d_n >$ *containing every element of D.*

**Output:** *A set of formulas EXTENSION and a set of default rules* $S \subseteq D$.

$S = \{d_1\}$; $i = 1$

*do while NOT(INTERRUPT) and* $S \subset D$

$S = S \cup \{d_{i+1}\}$; $i = i + 1$.

*EXTENSION=COMPUTE-EXTENSION*$((W, S), < d_1, d_2, \ldots, d_m >)$

*where* $< d_1, d_2, \ldots, d_m >)$ *is any arbitrary permutation of the elements of S* *return* $< EXTENSION, S >$

*stop*

**Algorithm 5** *BETA-SET-MEMBERSHIP*

**Input:** *A default theory (W,D), a logical variable INTERRUPT, a formula F and a sequence* $< d_1, d_2, \ldots, d_n >$ *containing every element of D.*

**Output:** *A logical variable IN-EXTENSION and a set of default rules* $S \subseteq \lambda((W, D))$.

$S = \{d_1\}$; $i = 1$

*do while NOT(INTERRUPT) and* $S \subset D$

$S = S \cup \{d_{i+1}\}$;$i = i + 1$

*for every permutation* $< d_1, d_2, \ldots, d_m >$) *of the elements of S*

*EXTENSION=COMPUTE-EXTENSION*($\mathcal{R}((W, D), S), < d_1, d_2, \ldots, d_m >$)

*if* $\{\neg F\}\cup$*EXTENSION is unsatisfiable then*

*return* $< IN - EXTENSION{=}TRUE, S >$

*stop*

*return* $< IN - EXTENSION{=}FALSE, S >$

*stop*

**Algorithm 6** *BETA-SET-ENTAILMENT*

**Input:** *A default theory (W,D), a logical variable INTERRUPT, a formula F and a sequence* $< d_1, d_2, \ldots, d_n >$ *containing every element of D.*

**Output:** *A logical variable IN-ALL-EXTENSIONS and a set of default rules* $S \subseteq D$.

$S = \{d_1\}$; $i = 1$

*do while NOT(INTERRUPT) and* $S \subset D$

$S = S \cup \{d_{i+1}\}$;$i = i + 1$

*IN-ALL-EXTENSIONS=TRUE*

*for every permutation* $< d_1, d_2, \ldots, d_m >$) *of the elements of S*

*EXTENSION=COMPUTE-EXTENSION*($(W, D), < d_1, d_2, \ldots, d_m >$)

*if* $\{\neg F\} \cup EXTENSION$ *is satisfiable then*

*IN-ALL-EXTENSIONS=FALSE*

*return* $< IN - ALL - EXTENSIONS, S >$

*stop*

We can state some fairly obvious properties of the above algorithms. We shall use the term *arbitrary termination* to denote the setting of the input variable INTERRUPT to TRUE at any point between start and completion of an anytime algorithm. $E(W, D)$ stands for the set of extensions of the default theory $(W, D)$.

**Observation:**

1. Let $< EXTENSION, S >$ be the output for some arbitrary termination of the algorithm ALPHA-COHERENCE, given a default theory $(W, D)$ and a sequence $< l_1, l_2, \ldots, l_n >$ composed of every element of $\lambda((W, D))$. Then $EXTENSION \in E(\mathcal{R}((W, D), S))$.

2. Let $< IN - EXTENSION, S >$ be the output for some arbitrary termination of the algorithm ALPHA-SET-MEMBERSHIP, given a default theory $(W, D)$, a formula F and a sequence $< l_1, l_2, \ldots, l_n >$ composed of every element of $\lambda((W, D))$. Then $\exists e : (e \in E(\mathcal{R}((W, D), S))) \land (e \models F)$.

3. Let $< IN - ALL - EXTENSIONS, S >$ be the output for some arbitrary termination of the algorithm ALPHA-SET-ENTAILMENT, given a default theory $(W, D)$, a formula F and a sequence $< l_1, l_2, \ldots, l_n >$ composed of every element of $\lambda((W, D))$. Then $\forall e : (e \in E(\mathcal{R}((W, D), S)))$ implies $e \models F$.

4. Let $< EXTENSION, S >$ be the output for some arbitrary termination of the algorithm BETA-COHERENCE, given a default theory $(W, D)$ and a sequence $< d_1, d_2, \ldots, d_n >$ composed of every element of D. Then $EXTENSION \in E(W, S)$.

5. Let $< IN - EXTENSION, S >$ be the output for some arbitrary termination of the algorithm BETA-SET-MEMBERSHIP, given a default theory $(W, D)$, a

formula F and a sequence $< d_1, d_2, \ldots, d_n >$ composed of every element of D. Then $\exists e : (e \in E(W, S) \wedge (e \models F)$.

6. Let $< IN - ALL - EXTENSIONS, S >$ be the output for some arbitrary termination of the algorithm BETA-SET-ENTAILMENT, given a default theory $(W, D)$, a formula F and a sequence $< d_1, d_2, \ldots, d_n >$ composed of every element of D. Then $\forall e : (e \in E(W, S)$ implies $e \models F$.

## 5.4 Anytime strategies based on PCS techniques

In this section, we shall attempt to delineate the space of possible anytime procedures for default inference that may be based on PCS techniques. Two observations are of interest here:

- PCS techniques based on breadth-first backtrack search correspond directly to procedures for computing $\alpha$-partial solutions. Progressively larger subsets of the set of variables (and hence, via our translation, of the set of propositional letters) are considered in progressively deeper breadth-first passes through the search tree. Arbitrary termination of such procedures will return solutions which are correct with respect to the set of variables (and hence, propositional letters) considered up to the most recent complete breadth-first pass through the search tree upto a certain depth.
- PCS techniques based on heuristic repair techniques, in the sense of [36], correspond directly to procedures for computing $\beta$-partial solutions. An initial assignment of values to variables (which must be consistent with essential constraints for W-preserving solutions) is repaired by progressively considering larger subsets of the set of potential constraints (and hence, progressively larger subsets of the default rules).

This does not preclude the possibility of using heuristic repair techniques for $\alpha$-partial solutions or using breadth-first backtrack search for $\beta$-partial solutions. The correspondence would not be as direct, however.

For brevity, we shall present only an instance of the first class of algorithms in detail.

In the following, we define an algorithm for returning $\alpha$-partial solutions to default inference problems. We define a single procedure for answering the three different kinds of queries by providing as input a variable QUERY which can take on any value from {COHERENCE, MEMBERSHIP, ENTAILMENT}. The algorithm calls a procedure PCS-BFS (for partial constraint satisfaction breadth first search). *INTERRUPT* operates as a global variable that the procedure PCS-BFS refers to. We assume that the variable *SearchFrontier* contains a set of pairs, where the first element of the pair is a partial solution (i.e., a set of variable assignments) and the second element is a set of constraints that this partial solution violates. We also assume that a procedure PCSP-Translate exists for translating a default theory into a PCSP specification. We assume that the priority relation on the constraints so generated, $\preceq$ partitions the set of constraints into two classes: $C_W$, the set of constraints obtained from the elements of $W$ and $C_D$, the set of constraints obtained from the elements of $D$. *MaxConstraints* is the set of subsets of $C_D$ that correspond to a maximal solution at any point. As before, the input variable sequence can be used to represent a priority relation on the set of propositional letters in the default theory. The variable QC (for query constraint) contains the translation of the formula $\neg F$ into a constraint in the same sense as described earlier.

**Algorithm 7** *ALPHA-PCS*
**Input:** *A default theory* $(W, D)$*, a formula F, a logical variable INTERRUPT, a variable QUERY, a sequence* $< v_1, v_2, \ldots, v_n >$ *containing every element of* $\lambda((W, D))$.
**Output:** *A 4-tuple < EXTENSION, MEMBER, ENTAILED, VARIABLES >. EXTENSION contains some element of E(W,D) in case QUERY=COHERENCE and NIL otherwise. MEMBER=TRUE in case QUERY=MEMBERSHIP and* $\exists e : (e \in E(W, D)) \wedge (e \models F)$*. Otherwise, MEMBER=FALSE. ENTAILED=TRUE in case QUERY=ENTAILMENT and* $\forall e : e \in E(W, D) \rightarrow e \models F$*. Otherwise, ENTAILED=FALSE.*

*VARIABLES contains the set of propositional letters considered in computing the current solution.*

*PCSP=PCS-Translate((W,D))*

*QC=PCSP-Translate*$(\{\neg F\}, \{\})$)

*PCS-BFS*$(\{< \{\}, \{\} >\}, \{\}, < v_1, v_2, \ldots, v_n >, \{\},$ *QUERY, PCSP)*

*stop*

*procedure PCS-BFS(SearchFrontier, MaxConstraints, VariablesRemaining, VariablesDone, QUERY, PCSP, QC)*

*If VariablesRemaining is the empty sequence then*

*return* $< MaxConstraints, FALSE, FALSE, VariablesDone >$

*stop*

*else*

*for NewAssignment=*$v_1 = TRUE, v_1 = FALSE$, *where* $v_1$ *is the first element of the VariablesRemaining sequence*

*for each* $< Soln, ViolatedConstraints > \in SearchFrontier$

*If INTERRUPT then*

*return* $< MaxConstraints, FALSE, FALSE, VariablesDone >$

*stop*

*else*

*NewSoln= Soln* $\cup$ *NewAssignment*

*If NewSoln violates no element of* $C_W$ *then*

*ViolatedConstraints=subset of* $C_D$ *that NewSoln violates*

*NewSearchFrontier=NewSearchFrontier* $\cup < NewSoln, ViolatedConstraints >$

*MaximalSolns= NewSearchFrontier* $-\{< S, V > | \exists < S', V' > \in NewSearchFrontier$ *s.t.* $V' \subset V\}$

*MaxConstraints=*$\{c \mid c \subseteq C_D$, *there exists some* $s \in MaximalSolns$ *s.t.* $s$ *satisfies* $c$ *and there is no* $c'$ *s.t.* $c \subset c' \subseteq C_D$ *s.t.* $s$ *satisfies* $c'\}$

*If QUERY=MEMBERSHIP and there exists*

*< Soln, ViolatedConstraints >∈ MaximalSolns s.t. Soln violates QC then*

*return < MaxConstraints, TRUE, FALSE, VariablesDone* $\cup\{v_1\}$ >

*stop*

*elseif QUERY=ENTAILMENT and for each*

*< Soln, ViolatedConstraints >∈ MaximalSolns, Soln violates QC then*

*return < MaxConstraints, FALSE, TRUE, VariablesDone* $\cup\{v_1\}$ >

*stop*

*else*

*PCS-BFS(NewSearchFrontier, MaxConstraints,* $< v_2, v_3, \ldots, v_n >$, *VariablesDone* $\cup\{v_1\}$,

*QUERY, PCSP, QC)*

*stop*

Notice that the algorithm above implicitly realizes a weak restriction function for default theories.

**Observation:** Let < MaxConstraints, MEMBER, ENTAILED, VariablesDone > be the output for some arbitrary termination of algorithm ALPHA-PCS, given a default theory $(W, D)$.

- $E(\mathcal{R}((W,D),S)) = \{Cn(W \cup d) \mid$ d is the subset of D corresponding to some element $c \in$ MaxConstraints$\}$.
- If QUERY=MEMBERSHIP, then MEMBER=TRUE iff $\exists e \in E(\mathcal{R}((\mathcal{W},\mathcal{D}),\mathcal{S}))$ s.t. $e \models F$ where $F$ is the query formula.
- If QUERY=ENTAILMENT, then ENTAILED=TRUE iff $\forall e : (e \in E(\mathcal{R}((\mathcal{W},\mathcal{D}),\mathcal{S}))) \rightarrow$ $(\rceil \models \mathcal{F})$, where $F$ is the query formula.

# Chapter 6

# Conclusion

## 6.1 Contributions

In this dissertation, we seek to provide a framework for the design of practical systems for belief change. We do this through the following steps:

*Competence:* The work of Alchourrón, Gärdenfors and Makinson [1], [19], provides a comprehensive and widely accepted competence theory for the process of belief change. We identify the following major drawbacks in this theory:

- It provides an inadequate account of the process of retracting a belief. Thus, the addition of a belief is duly recorded in the belief state of an agent, but the retraction of a belief is never recorded. This can unduly restrict the space of candidate outcomes of a belief change operation.
- The theory provides no prescription on how beliefs must change when the belief input is not fully credible. Any approach to handling uncertain, or less credible, belief inputs should involve a generalization of techniques applied when the belief inputs are fully credible, instead of requiring a totally distinct set of techniques.
- It is generally agreed that the *principle of informational economy* should guide any strategy for belief change. This requires that beliefs should be

discarded as little as possible while effecting belief change. The competence theory of Alchourrón, Gärdenfors and Makinson seeks to satisfy this requirement, but with limited success. As a consequence of the belief representation scheme and an unduly narrow definition of what constitutes success for a belief change operation, beliefs may be unduly discarded by operators defined within this framework.

- The theory does not specify belief change beyond a single step. Several authors, such as [38], [37] and citeDarwiche94 have sought to address this question, but their solutions suffer from the previous three problems.

We develop a theory that accounts for each of the problems mentioned above, and argue that it provides an adequate set of benchmark tests, as well as a suitable starting point for implemented belief change systems.

*Performance:* We present the design of two belief change systems which use a variant of default logic [14] as the belief representation language. The design of the first system preceded the development of the our competence theory and provided the motivation for this theory, by identifying several of the lacunae in the existing definition of competence. The second system was developed using our competence theory as the starting point. These two designs serve to demonstrate that practically implementable systems that satisfy the requirements identified in our competence theory are indeed possible. The use of a default logic variant has several other practical benefits as well, such as the ability to incorporate lazy evaluation strategies in computing belief change.

*Implementation:* Belief change is a computationally hard problem [40], including our formulation of the problem in the two systems mentioned above. Nevertheless, practical constraints often require tractable solutions, or procedures that exhibit resource-bounded rationality [48]. We present a toolkit of two approaches to address such concerns. First, we define a mapping from the problem of default inference to partial constraint satisfaction problems [18]. The mapping enables us to apply techniques from the area of partial constraint satisfaction to improve the efficiency of procedures for computing default extensions, and hence for computing belief change. Next, we present a set of strategies for computing meaningful partial results in resource-bounded situations, by defining *anytime* procedures for default inference. While much remains to be done in this area, we believe these strategies can provide the basis for fielded applications of problem solvers with a significant belief change component.

## 6.2 Future Work

### 6.2.1 Reasoning about Static vs. Dynamic Worlds

*Revision* vs. *Update*

The close relationship between the problem of belief revision and the problem of reasoning about action is a well-recognized one [17], [52], [29], [13]. Both problems involve the question of how to update a set of beliefs (in the case of reasoning about action, these beliefs denote our knowledge about the state of the world), in order to accommodate the effects of some change. The following example is a toy problem which shows how the machinery of belief revision can be used to reason about changes in the world as a result of some action.

**Example:** Let the initial state of the world, to an agent whose knowledge is limited to the weather system in a city, be one in which it is not snowing. Let the agent's knowledge-base of the physical laws that govern this weather system consist of a rule that indicates that if it is snowing, it must be cold. The initial belief state of the agent is thus given by the theory $(W, D)$ where $W = \{snowing \rightarrow cold, \neg snowing\}$ and $D = \{\frac{:\top}{\top}\}$. To identify the new state of the world after it starts snowing, we add the belief *snowing* to this belief state. The new belief state is given by the theory $(W', D')$ where $W' = \{snowing, snowing \rightarrow cold\}$ and $D' = \{\frac{:\neg snowing}{\neg snowing}\}$. The new PJ-default theory has exactly one extension which correctly indicates that it must be cold now. □

Unlike this example, however, the techniques of belief revision are not always directly applicable to reasoning about dynamic worlds. Katsuno and Mendelzon [29] identify two distinct kinds of belief change: *revision* and *update* (in the rest of this report, we shall italicize the two terms whenever we use them in the sense of Katsuno and Mendelzon). *Revision* is used to denote the operation of modifying a belief state in the light of new information about a static world, whereas *update* denotes the operation of modifying a belief state to incorporate new information indicating that the world has changed as a result of some action. They explain this distinction using a model-theoretic argument. In *revising* a knowledge-base $\phi$ with a new sentence $\mu$, *revision* methods that satisfy the AGM postulates select from the models of $\mu$, those that are "closest" to models of $\phi$, where closeness is defined by an ordering relation amongst models that satisfies certain conditions. The selected models determine the *revised* theory. In *updating* a knowledge-base $\phi$ with a new sentence $\mu$, for each model $M$ of $\phi$, the set of models of $\mu$ that are closest to $M$ are selected. The resultant knowledge-base is determined by the union of the selected models. Another way to view this distinction is to treat *revision* as the problem of changing our body of knowledge of the world at time point $t$, given new information about the state of the world at time point $t$, and *update* as the problem of deciding which facts about the world at time point $t$ continue to be true at time point $t+1$, given new information about the world at $t+1$. In this context, the issue of persistence of beliefs appears to be closely related to the notion of persistence in theories of action. The following variation of the previous example shows how our framework defined earlier in this report is not directly applicable in cases of *update*.

**Example:** Let the initial state of the world be denoted, as before, by $(W_0, D_0)$ where $W_0 = \{snowing \rightarrow cold, \neg snowing\}$ and $D_0 = \{\frac{:\top}{\top}\}$. Let us now *update* the belief state with the belief that it is not cold. The new belief state is given by $(W_1, D_1)$ where $W_1 = \{snowing \rightarrow cold, \neg snowing, \neg cold\}$ and $D_1 = \{\frac{:\top}{\top}\}$. Now, let us *update* this belief state with the belief that it is snowing. As a consequence of prioritizing the belief constraints in the temporal order in which they are added, the resultant belief state $(W_2, D_2)$ is given by $W_2 = \{\neg cold, snowing\}$ and $D_2 = \{\frac{:snowing \rightarrow cold}{snowing \rightarrow cold}, \frac{:\neg snowing}{\neg snowing}\}$. Notice that the resultant belief state contains the unintuitive belief that it is still not cold, inspite of the fact that it is now snowing. The unintuitive result is a consequence of our assigning to the rule *snowing* $\rightarrow$ *cold* the status of a defeasible belief, inspite of it being a physical law that is never violated. □

It turns out that prioritizing belief constraints in the temporal order of their arrival is suitable only in cases of *revision.* A belief constraint set consisting of three belief constraints prioritized as $a \rightarrow b \prec -b \prec a$ in the temporal order of their arrival will produce reasonable results only if we have a guarantee that the world has remained static over these three belief change steps. We are first told that $a \rightarrow b$ is true, then told to disbelieve $b$ and finally told to believe $a$. Given that the world has not changed, it appears reasonable to have more faith in the more recent evidence and disbelieve $a \rightarrow b$.

It must be recognized that in real life it is often difficult to distinguish between *revisions* and *updates.* Given a new piece of information, it might be difficult to decide whether this information is a consequence of a changed world or whether this is merely new information about the same world. However, given that the two cases must be treated in entirely different ways, the distinction should be utilized wherever it can be identified.

### *Updating* Belief States in Dynamic Worlds

The previous example suggests a different constraint prioritization protocol that might provide more intuitive results in the case of *updates.* The new approach involves assigning a special status to those belief constraints which denote physical laws or inviolate facts about the world, in a manner similar to the status accorded to database integrity constraints in database updates. Assuming this corpus of physical laws and inviolate facts to be mutually consistent, we shall reserve the highest priority class for this set of belief constraints. Other belief constraints are prioritized as before, in the temporal order of their arrival. A similar technique, involving assigning the highest epistemic entrenchment in the AGM framework to physical laws, has been used by Foo and Rao [17] in a planning framework that uses AGM belief change operators. The following example shows how this new prioritization protocol avoids the unintuitive results of the previous example.

**Example:** We assign to *snowing* $\rightarrow$ *cold* the status of a physical law. At the third belief change step in the previous example, the belief constraints will now be prioritized as follows: $\neg snowing \prec \neg cold \prec snowing \prec snowing \rightarrow cold$. The final belief state $(W_2, D_2)$ will thus be given by $W_2 = \{snowing \rightarrow cold, snowing\}$ and $D_2 = \{\frac{:\neg cold}{\neg cold}, \frac{:\neg snowing}{\neg snowing}\}$. The PJ-default theory has a single extension which correctly indicates that it is snowing and that it is cold. □

The important lesson here is that the belief constraint prioritization is a crucial parameter that may need to be altered to make a belief change framework suitable for different classes of applications. More generally, the role of evidence is an important factor that varies across classes of applications for a belief change framework.

However, a new constraint prioritization protocol turns out to be insufficient for obtaining a belief change framework that meets all the needs of the *update* operation. We shall motivate the discussion with an example based on [29] which shows how a syntax-based approach similar to the framework defined in this report fails to provide intuitive results in some situations while fully rational *update* operators (in the sense of a set of rationality postulates for *update* defined by Katsuno and Mendelzon [29]) do provide intuitive results. We briefly summarize the Katsuno and Mendelzon postulates for *update* below. In the following, the result of *updating* a knowledge-base $\phi$ with the sentence $\mu$ is denoted by $\phi \diamond \mu$.

(U1) $\phi \diamond \mu$ implies $\mu$.

(U2) If $\phi$ implies $\mu$, then $\phi \diamond \mu$ is equivalent to $\phi$.

(U3) If both $\phi$ and $\mu$ are satisfiable, then $\phi \diamond \mu$ is also satisfiable.

(U4) If $\phi_1 \equiv \phi_2$ and $\mu_1 \equiv \mu_2$ then $\phi_1 \diamond \mu_1 \equiv \phi_2 \diamond \mu_2$.

(U5) $(\phi \diamond \mu) \wedge \psi$ implies $\phi \diamond (\mu \wedge \psi)$.

(U6) If $\phi \diamond \mu_1$ implies $\mu_2$ and $\phi \diamond \mu_2$ implies $\mu_1$, then $\phi \diamond \mu_1 \equiv \phi \diamond \mu_2$.

(U7) If $\phi$ is complete, then $(\phi \diamond \mu_1) \wedge (\phi \diamond \mu_2)$ implies $\phi \diamond (\mu_1 \vee \mu_2)$.

(U8) $(\phi_1 \vee \phi_2) \diamond \mu \equiv (\phi_1 \diamond \mu) \vee (\phi_2 \diamond \mu)$.

Postulates (U1)-(U5) are equivalent to a subset of the AGM postulates for revision reformulated by Katsuno and Mendelzon in a model-theoretic framework. Of special interest is postulate (U8), the so-called "disjunction rule" which requires that each possible model of the current state of the world $\phi$ should be separately updated with the new sentence $\mu$ to obtain $\phi \diamond \mu$. This is required even if $\mu$ is consistent with $\phi$, unlike *revision*, which maps to the new belief state $\phi \wedge \mu$.

For the example below, we shall use Winslett's [52] Possible Models Approach (or PMA) operator, which has been shown to be a rational *update* operator in the sense of the postulates presented above. In *updating* belief state $\phi$ with $\mu$, the PMA operator selects, for each model $I$ of $\phi$, those models of $\mu$ which are "closest" to $I$. The union of all such models is taken to denote $\phi \diamond \mu$. Formally:

$$Mod(\phi \diamond \mu) = \bigcup_{I \in Mod(\phi)} Incorporate(Mod(\mu), I)$$

where $Mod(\psi)$ denotes the set of models of $\psi$ and $Incorporate(Mod(\mu), I)$ stands for the set of models of $\mu$ that are "closest" to $I$. Closeness between two models $I$ and $J$ is measured by the set $Diff(I, J)$ of propositional letters that have different truth values in $I$ and $J$. A model $J_1$ is closer to $I$ than a model $J_2$ (written $J_1 \leq_I J_2$) iff $Diff(I, J_1) \subsetneq Diff(I, J_2)$. $Incorporate(Mod(\mu), I)$ consists of the minimal elements of $Mod(\mu)$ with respect to $\leq_I$.

**Example:** Let our propositional language consist of only two letters: $b$ and $m$. Let $b$ denote that the book is on the floor, and let $m$ denote that the magazine is on the floor. The initial state of the world (in this case, a room we cannot see), about which the only information available is that either the book or the magazine is on the floor but not both, is given by $\phi = (b \wedge \neg m) \vee (\neg b \wedge m)$. We now order a robot to enter the room and place the book on the floor; in other words, $\mu = b$. Assume that we have a guarantee that moving the book will have to effect on the position of the magazine.

In System *BR1*, the initial belief state is given by the theory $(W, D)$ where $W = \{(b \wedge \neg m) \vee (\neg b \wedge m)\}$ and $D = \{\frac{:\top}{\top}\}$. Since $\mu$ is consistent with the contents of $W$, the *updated* belief state $(W', D')$ is given by $W' = \{((b \wedge \neg m) \vee (\neg b \wedge m)) \wedge b\}$ and $D' = \{\frac{:\top}{\top}\}$. The conjunction of the contents of $W'$ is equivalent to the sentence $b \wedge \neg m$. In others words, we conclude, in the *updated* belief state that the book is on the floor and the magazine is not. While the conclusion that the book is on the floor is reasonable, the conclusion that the magazine is not is unwarranted. Such an outcome commits us to the view that the previous state of the world was one in which the book was on the floor while the magazine was not (since the only thing that could possibly be affected by the robot's action is the position of the book). There is nothing in the evidence, however, that indicates that this should have been the case. In fact, the only thing that can be reasonably concluded about the new state of the world, given the available information, is that the book is on the floor. The outcome of the PMA operator is precisely this. $\phi$ has two models, $\{b, \neg m\}$ and $\{\neg b, m\}$. The model of $\mu$ closest to the first model is $\{b, \neg m\}$ while the model of $\mu$ that is closest to the second model is $\{b, m\}$. The outcome is given by the union of these two models, so that $\phi \diamond \mu = b$. □

Obviously, *update* operations require very different treatment in the context of Systems *BR1* and *BR2*. We provide below a preliminary proposal of a method to address this problem.

- Given an initial knowledge-base describing the state of the world, represented as a propositional theory, we compute its equivalent in disjunctive normal form (DNF).

- The initial belief state is denoted by the theory $(W, D)$ where $W$ is empty and $D$ consists of each disjunct of the DNF theory represented as a normal default (recall that a normal default is a rule of the form $\frac{:\phi}{\phi}$. Thus, the DNF theory

$\alpha \vee \beta \vee \gamma$ gets mapped to the set of PJ-default rules $D = \{\frac{:\alpha}{\alpha}, \frac{:\beta}{\beta}, \frac{:\gamma}{\gamma}\}$.

- To *update* such a PJ-default theory with a new belief $\mu$, we define an operator called UPDATE-WORLD which maps a set of PJ-default rules to another set of PJ-default rules.

$$\text{UPDATE-WORLD}(D, \mu) = \{\frac{:\alpha' \wedge \mu \wedge \Gamma}{\alpha' \wedge \mu} \mid \alpha' \in (\alpha \downarrow \neg\mu), \frac{:\alpha \wedge \Gamma}{\alpha} \in D\}$$

For each PJ-default rule of the form $\frac{:\alpha \wedge \Gamma}{\alpha}$ (where $\alpha$ is the "normal" part of the default rule and denotes a tentative belief in our framework, while $\Gamma$ is the "semi-nromal" part of the default rule and contains the necessary disbelief constraints in our framework), the UPDATE-WORLD operator identifies the maximal subsets of the set of conjuncts that constitute $\alpha$ (we assume that $\alpha$ is in conjunctive normal form) that are consistent with $\mu$, and for each such subset $\alpha'$, creates a new PJ-default rule of the form $\frac{:\alpha' \wedge \mu \wedge \Gamma}{\alpha' \wedge \mu}$. Thus if $(W', D') = (W, D) \diamond \mu$, then $W' = W$ and $D'$ =UPDATE-WORLD$(D, \mu)$.

The intuition here is that representing the current state of the world in DNF is a syntactic approximation of identifying the different possible views of the current world. To allow for the fact that different views of the world can be mutually inconsistent, we represent each of the disjuncts of the DNF as default rules in our representation of belief states as PJ-default theories. Each extension of the PJ-default theory thus represents a consistent picture of the world. In accordance with Katsuno and Mendelzon's [29] *disjunction rule*, we *update* each possible view of the world separately, by altering each default rule using the UPDATE-WORLD operator. *Revision*, in contrast, would require that we add the new belief to $W$. The following example shows how this approach might address the problem encountered in the previous example.

**Example:** The initial state of the world $\phi = (b \wedge \neg m) \vee (\neg b \wedge m)$, is represented as the theory $(W, D)$, where $W = \{\}$ and $D = \{\frac{:(b \wedge \neg m)}{(b \wedge \neg m)}, \frac{:(\neg b \wedge m)}{(\neg b \wedge m)}\}$. It has two extensions: $\{b \wedge \neg m\}$ and $\{\neg b \wedge m\}$. $(W', D') = (W, D) \diamond b$ is given by $W' = \{\}$ and $D' = \{\frac{:(b \wedge \neg m)}{(b \wedge \neg m)}, \frac{:(b \wedge m)}{(b \wedge m)}\}$. This has two extensions: $\{b \wedge \neg m\}$ and $\{b \wedge m\}$, which correspond precisely to the two models obtained as a result of applying the PMA operator. The only thing that a sceptical reasoner will be able to conclude about the new state of the world is that $b$ is true, which is exactly what intuition dictates. □

In the same way that *update* is defined to be a dynamic worlds counterpart of *revision*, Katsuno and Mendelzon define *erasure* to be the dynamic worlds counterpart of AGM contraction. They define the erasure of $\mu$ from $\phi$ to entail the addition of models to the set of models of $\phi$. For each model $I$ of $\phi$, those models in which $\mu$ is not true that are closest to $I$ are added. This definition of erasure, however, appears to be unintuitive, as the following example shows.

**Example:** As before, $\phi = (b \wedge \neg m) \vee (\neg b \wedge m)$ from which we shall erase $b$. The model $M_1 = \{\neg b, m\}$ of $\phi$ needs no further change since $b$ is not true in it. Model $M_2 = \{b, \neg m\}$ makes $b$ true, so we identify $M_3 = \{\neg b, \neg m\}$ as the model which does not make $b$ true and which is closest to $M_2$. According to Katsuno and Mendelzon's definition of erasure, we take the union of $M_1$, $M_2$ and $M_3$ to denote the new state of the world. This is equivalent to the formula $(b \wedge \neg m) \vee \neg b$. Such a result runs counter to intuition since it indicates that $b$ is true in one possible view of the new state of the world. A more reasonable result is obtained if the union of only $M_1$ and $M_3$ is taken. This is equivalent to the formula $\neg b$ which corresponds precisely to the intuition that after being told that the book is no longer on the floor, all that we can reasonably conclude about the world is just that. □

We therefore argue that the erasure operator is better formalized as follows. The erasure of $\mu$ from $\phi$ (written $\phi \bullet \mu$) is given by the union of those models which do not make $\mu$ true which are closest to $I$, for each $I$ that is a model of $\phi$. In terms of our framework, the steps would be exactly the same as those for *update*, except that in place of UPDATE-WORLD, the operator ERASE-WORLD, defined below, would be used.

$$\text{ERASE-WORLD}(D, \mu) = \{\tfrac{:\alpha' \wedge \neg\mu \wedge \Gamma}{\alpha' \wedge \neg\mu} \mid \alpha' \in (\alpha \downarrow \mu), \tfrac{:\alpha \wedge \Gamma}{\alpha} \in D\}$$

It is easy to see that in the previous example, with our framework, the new belief state would be PJ-default theory with two extensions: $\{\neg b \wedge m\}$ and $\{\neg b, \neg m\}$.

The applicability of a syntax-based frameworks such as Systems *BR1* and *BR2* to problems of *update* appears to be a promising direction of research. A variety of issues need to be addressed. The dynamics of repeated belief change operations using the UPDATE-WORLD and ERASE-WORLD operators needs to be examined. The rationality of these operators with respect to the *update* postulates listed above, as well as with respect to a similar set of postulates for erasure, needs to be investigated. A comprehensive framework for iterated belief change that offers the full repertoire of *update*, *revision*, *erasure* and *contraction* operators needs to be developed. Such an integrated framework can provide interesting insights; whereas *updates* treat physical laws to be inviolate, they can be doubted as result of *revisions* (this would correspond to a paradigm shift in science). Such studies should also shed light on the relationship between syntax-based and model-theoretic approaches to belief change. This research can be viewed as a first step toward making a theory of belief change applicable as theory of actions and change.

### 6.2.2 The Problem of Combining Knowledge Bases

The problem of combining knowledge-bases is a non-trivial one and has wide-ranging applications. The problem can arise when several experts contribute to the development of a knowledge-base system. A reasonable assumption is that the body of knowledge of each individual expert will be consistent, but no such guarantee exists for the sum total of all their inputs. The problem of combining knowledge-bases is thus the problem of handling the inconsistencies that may arise when several different, internally consistent knowledge-bases are joined together. Such questions are becoming increasingly relevant with the current trend towards very large knowledge bases [31], and towards reusable and shareable knowledge-bases [41]. The issues addressed in this problem domain are essentially the same as those addressed in belief revision. Combining two knowledge-bases can be viewed as *revising* one knowledge-base with another, although in some frameworks (such as ours), this could entail the unintended assignment of a higher priority to the second knowledge-base over the first. It is clear, however, that theories of belief change can provide a good basis for developing principled methods of combining knowledge-bases.

The study by Baral, Kraus, Minker and Subrahmaniam in [2] represents a first step in addressing this problem. They present four combination functions and attempt to identify situations where each function might be appropriate. We present the definitions of the functions $Comb_1, Comb_3$ and $Comb_4$ below ($Comb_2$ is ommitted since it is shown to be equivalent to $Comb_1$). Each individual knowledge-base is represented as a first-order theory, and each combination function involves a different approach to identifying maximal subsets of the union of these theories.

$Comb_1$ is defined as follows:

$$Comb_1(\{T_1, \ldots, T_n\}, IC) = MAXCONS(T_1 \cup \ldots \cup T_n, IC)$$

$Comb_1(\{T_1, \ldots, T_n\}, IC)$ denotes the combination of the theories $T_1, \ldots, T_n$ respecting a set of integrity constraints. $MAXCONS(P, IC)$ denotes the set of maximal consistent subsets of $P$ with priority to $IC$; in other words, $MAXCONS(P, IC)$ contains every $Q \subseteq P$ such that $Q \cup IC$ is consistent and for every $Q'$ such that $Q \subset Q' \subseteq P$, $Q' \cup IC$ is inconsistent. $Comb_3$ postpones the checking consistency with $IC$ until after the maximal consistent subsets of the union of the theories have been identified, as the following definition shows:

$$Comb_3(\{T_1, \ldots, T_n\}, IC) = \{X \mid X \in MAXCONS(T_1 \cup \ldots \cup T_n) \text{ and } X \cup IC \text{ is consistent }\}$$

In the definition above, $MAXCONS(P)$ stands for the restriction of $MAXCONS(P, IC)$ to the case where $IC = \emptyset$. Both $Comb_1$ and $Comb_3$ use maximality with respect to set inclusion to identify maximal consistent subsets, and provide reasonable results under two slightly different intuitions of what the combination process should involve. $Comb_4$, however, uses a somewhat suspect notion of maximality with respect to set cardinality, as the definition below shows:

$$Comb_4(\{T_1, \ldots, T_n\}, IC) = \text{maximal elements w.r.t. set cardinality of } Comb_1(\{T_1, \ldots, T_n\}, IC)$$

Default theories appear to be a reasonable framework for representing the outcome of combining mutually inconsistent knowledge-bases, specially given that their properties are well-understood and given that well-defined procedures exist for computing extensions. For instance, the function $Comb_1$ could be reformulated as a mapping from a set of theories to a PJ-default theory $(W, D)$ such that $W = IC$ and $D$ consists of the elements of the theories represented as normal default rules (i.e., every $\alpha \in T_i$ is represented as $\frac{:\alpha}{\alpha}$). Alternatively, one could trade-off "compile-time" efficiency for "run-time" efficiency by setting:

$$W = \bigcap MAXCONS(\{T_1, \ldots, T_n\}, IC)$$
$$D = \{\tfrac{:\alpha \wedge IC}{\alpha} \mid \alpha \in (E - W), E \in MAXCONS(\{T_1, \ldots, T_n\}, IC)\}$$

Thus, while computing $(W, D)$ will involve more effort, skeptical reasoning with the combined theory will involve no effort at all since $W$ will contain all the relevant knowledge.

A need exists for a study generalizing these results for the case combining non-monotonic theories. We believe that our framework provides a good starting point for such studies. Even more interesting possibilities exist when one considers the problem of combining knowledge-bases represented, like belief states in our framework, as 4-tuples of the form < necessary-beliefs, necessary-disbeliefs, tentative-beliefs, constraint-prioritization >.

# Bibliography

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] Chitta Baral, Sarit Kraus, Jack Minker, and V. Subrahmaniam. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1):45–71, 1992.

[3] Alex Borgida. Language features for flexible handling of exceptions in information systems. *ACM Trans. on Database Systems*, 10:563–603, 1985.

[4] Craig Boutilier. Conditional logics for default reasoning and belief revision. Technical Report Technical Report 92-1, University of British Columbia, Department of Computer Science, 1992.

[5] Craig Boutilier. Sequences of revisions: On the semantics of nested conditionals. Technical Report Technical Report 92-23, University of British Columbia, Department of Computer Science, 1992.

[6] Gerhard Brewka. Belief revision in a framework for default reasoning. In *Proc. of the Konstanz Workshop on Belief Revision*, pages 206–222, 1989.

[7] Gerhard Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proc. of the 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[8] Marco Cadoli and Marco Schaerf. Approximate inference in default logic and circumscription. In *Proc. of the Tenth European Conf. on AI*, pages 319–323, 1992.

[9] Mukesh Dalal. Investigations into a theory of knowledge base revision : Preliminary Report. In *Proc. of the 7th National(USA) Conf. on Artificial Intelligence*, pages 475–479, 1988.

[10] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. In *Proc. of the 5th Conference on Theoretical Aspects of Reasoning about KNowledge*, pages 5–23, 1994.

[11] Johan de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28:127–162, 1986.

[12] T. Dean and M. Boddy. An analysis of time-dependent planning. In *Proc. of the Seventh National Conf. on AI*, pages 49–54, 1988.

[13] Alvaro del Val and Yoav Shoham. Deriving properties of belief update from theories of action. In *Proceedings of the Tenth Conference of AAAI*, 1992.

[14] J. P. Delgrande and W. K. Jackson. Default logic revisited. In *Proc. of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 118–127, 1991.

[15] Jon Doyle. A truth maintenance system. *Artificial Intelligence*, 12:231–272, 1979.

[16] Ronald Fagin, Jeffrey D. Ullman, and Moshe Y. Vardi. On the semantics of updates in databases. In *Proc. of the Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 352–365, 1983.

[17] Norman Y. Foo and Anand S. Rao. Belief revision in a microworld. Technical Report Tehcnical Report 325, Univ. of Sydney, Basser Dept. of Computer Science, 1988.

[18] Eugene C. Freuder and Richard J. Wallace. Partial constraint staisfaction. *Artificial Intelligence*, 58(1-3):21–70, 1992.

[19] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, MA, 1988.

[20] Peter Gärdenfors. Nonmonotonic inferences based on expectations: A preliminary report. In *Proc. of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 585–590, 1991.

[21] Peter Gärdenfors and David Makinson. Revisions of knowledge systems using epistemic entrenchment. In *Proc. of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 83–95, 1988.

[22] Peter Gärdenfors and Hans Rott. Belief revision. Technical Report 30, Zentrum Philosophie und Wissenschaftstheorie, Universität Konstanz, 1992.

[23] Aditya K. Ghose and Randy Goebel. Default reasoning as belief change: A rationale. In *Proc. of the 2nd Pacific Rim Int'l Conf. on AI*, 1992.

[24] Aditya K. Ghose, Pablo O. Hadjinian, Abdul Sattar, Jia-H. You, and Randy Goebel. Iterated belief change. Technical Report TR93-05, University of Alberta, Dept. of Computing Science, 1993. Also published as Griffith University, School of Computing and Information Technology Research Report No. 62.

[25] Aditya K. Ghose, Pablo O. Hadjinian, Abdul Sattar, Jia-H. You, and Randy Goebel. Iterated belief change: A preliminary report. In *Proc. of the Sixth Australian Joint Conference on AI*, 1993.

[26] Aditya K. Ghose, Abdul Sattar, and Randy Goebel. Default reasoning as partial constraint satisfaction. In *Proc. of the Sixth Australian Joint Conference on AI*, 1993.

[27] Sven Ove Hansson. *Belief Base Dynamics.* Uppsala, 1991.

[28] W. L. Harper. Rational belief change, Popper functions and counterfactuals. *Synthese*, (30):221–262, 1975.

[29] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proc. of the Second Int'l Conf. on Principles of Knowledge Representation and Reasoning*, 1991.

[30] H.A. Kautz and B. Selman. Hard problems for simple default logics. In *Proc. of the First International Conference on the Principles of Knowledge Representation and Reasoning*, pages 189–197, 1989.

[31] Douglas B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representations and Inference in the Cyc Project.* Addison-Wesley, Reading, MA, 1990.

[32] Isaac Levi. Subjunctives, dispositions and chances. *Synthese*, 34:423–455, 1977.

[33] Dekang Lin and Randy Goebel. Computing circumscription of ground theories with Theorist. Technical Report TR89-26, Department of Computing Science, University of Alberta, Edmonton, Canada, October 1989.

[34] David Makinson and Peter Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In *Proc. of the Konstanz Workshop on Belief Revision*, pages 185–205, 1989.

[35] J. Martins and S. Shapiro. A model for belief revision. *Artificial Intelligence*, 35(1):25–79, 1988.

[36] Steven Minton, Mark D. Johnston, Andrew B. Philips, and Philip Laird. Minimizing conflicts: a heuristic repair method for constraint satisfaction and scheduling problems. *Artificial Intelligence*, 58(1-3):161–205, 1992.

[37] A. C. Nayak, N. Y. Foo, M. Pagnucco, and A. Sattar. Entrenchment Kinematics 101. In *Proc. of the 7th Australian Joint Conference on Artificial Intelligence*, pages 301–311, 1994.

[38] Abhaya C. Nayak. *Studies in Belief Change.* Ph.D. thesis, Department of Philosophy, University of Rochester, 1993.

[39] Bernhard Nebel. A knowledge level analysis of belief revision. In *Proc. of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 301–311, 1989.

[40] Bernhard Nebel. Belief revision and default reasoning: Syntax-based approaches. In *Proc. of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 417–428, 1991.

[41] Robert Neches, Richard E. Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, and William R. Swartout. Enabling technology for knowledge sharing. *AI Magazine*, 12(3):36–56, 1991.

[42] David Poole, Randy Goebel, and Romas Aleliunas. Theorist: a logical reasoning system for defaults and diagnosis. In N.J. Cercone and G. McCalla, editors, *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 331–352. Springer Verlag, 1987.

[43] Teodor C. Przymusinski. An algorithm to computer circumscription. *Artificial Intelligence*, 38(1):49–73, 1989.

[44] F. P. Ramsey. *Foundations of Mathematics and Other Logical Essays.* Routledge and Kegan Paul, New York, NY, 1931.

[45] Anand S. Rao and Norman Y. Foo. Formal theories of belief revision. In *Proc. of the First International Conference on Principles of Knowledge Representation and Reasoning*, 1989.

[46] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1 and 2):81–132, 1980.

[47] Ken Satoh. Nonmonotonic reasoning by minimal belief revision. In *Proc. of the Int'l Conf. on Fifth Generation Computer Systems*, pages 455–462, 1988.

[48] Herbert A. Simon. A behavioural model of rational choice. *Quarterly Journal of Economics*, 69:99–118, 1955.

[49] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change and Statistics*, volume 2, pages 105–134. 1988.

[50] Andreas Weber. Updating propositional formulas. In *Proc. of the 1st Conf. on Expert Database Systems*, pages 487–500, 1986.

[51] M. A. Williams. Transmutations for theory bases. In *Proc. of the 6th Australian Joint Conference on AI*, pages 83–89, 1993.

[52] Marianne Winslett. Reasoning about action using a possible models approach. In *Proc. of the 7th National(USA) Conf. on Artificial Intelligence*, pages 89–93, 1988.

[53] L. Yuan and J. You. Knowledge base revision using circumscription. In *Proc. of the 3rd Int'l Conf. on Deductive and Object-Oriented Databases*, pages 44–458, 1993.