Belief Revision in Non-Monotonic Reasoning

José Alferes[⋆]
DM, U. Évora and CRIA Uninova
2825 Monte da Caparica, Portugal
(jja@fct.unl.pt)

Luís Moniz Pereira[∗]
DCS, U.Nova de Lisboa and CRIA Uninova
2825 Monte da Caparica, Portugal
(lmp@fct.unl.pt)

Teodor C. Przymusinski[⋆⋆]
Department of Computer Science
University of California
Riverside, CA 92521, USA
(teodor@cs.ucr.edu)

1 Introduction

Moore's autoepistemic logic, AEL [Moo85], was obtained by augmenting classical propositional logic with a modal operator $\mathcal{L}$. The intended meaning of the modal atom $\mathcal{L}F$ in the stable autoepistemic expansion T is "F is known" in T, or, more precisely, "F is logically derivable" from T . Indeed, a formula $\mathcal{L}F$ belongs to the expansion T if and only if F is derivable from T. Thus Moore's modal operator $\mathcal{L}$ can be viewed as a "knowledge operator" which allows us to reason about formulae *known* to be true in the expansion[3]. However, often times we need to reason about formulae that are only *believed* (rather than known) to be true, where what is believed or not believed is determined by some specific *non-monotonic formalism*. In particular, we may want to express beliefs based on *minimal entailment*, or, more generally, on some form of *circumscription*, and thus we may need a modal "belief operator" $\mathcal{B}$ with the intended meaning of $\mathcal{B}F$ given by "F is true in all minimal models" or "F is minimally entailed" (in the expansion).

In order to be able to explicitly reason about beliefs, in [Prz94b, Prz94a] the third author introduced a new non-monotonic formalism, called the *Autoepistemic Logic of Beliefs*, AEB, obtained by augmenting classical propositional logic with a *belief operator*, $\mathcal{B}$. The resulting non-monotonic knowledge representation framework turned out to be rather simple and yet quite powerful. It was proved that propositional circumscription and several major semantics for

⋆ Partially supported by JNICT-Portugal and ESPRIT project Compulog 2 (no. 6810).
⋆⋆ Partially supported by the National Science Foundation grant #IRI-9313061.
3 This view of Moore's modal operator $\mathcal{L}$ as a knowledge operator is not shared by all researchers. Moore himself considered $\mathcal{L}$ to be a belief operator.

logic programs are *isomorphically embeddable* into AEB. In particular, this is true for the stable, well-founded, stationary and static semantics for normal, disjunctive and extended logic programs [GL88, VGRS90, Prz91b, GL90, Prz94c].

At the same time the Autoepistemic Logic of Beliefs, AEB, has some very natural properties which sharply contrast with those of Moore's AEL. In particular, *every* belief theory T in AEB has the *least* (in the sense of inclusion) static expansion $T^{\diamond}$ which has an *iterative* definition as the *least fixed point* of a monotonic belief closure operator. Moreover, least static expansions are always consistent in the broad class of *affirmative* belief theories defined below.

While static expansions seem to provide a natural and intuitive semantics for many belief theories, and, in particular, for all normal and disjunctive logic programs, they often lead to inconsistent static expansions for theories in which (subjective) beliefs clash with the known (objective) information or with some other beliefs. In particular, this applies to extended logic programs with strong or explicit negation.

Consider the following sentence of AEB:

$$\mathcal{B}\neg FlatTire \wedge \mathcal{B}\neg BadBattery \supset \neg Broken$$

which is intended to say that in the absence of any indication that something is wrong with the tires or with the battery we can conclude that the car is not broken[4]. Assuming this is all we know about the car, we are likely to conclude that it is not broken because we have no indication that would make us believe that there is any problem with either battery or tires. In other words, both $FlatTire$ and $BadBattery$ are false in all minimal models of our knowledge base and thus both $\mathcal{B}\neg FlatTire$ and $\mathcal{B}\neg BadBattery$ hold true.

Suppose, however, that upon inspection we learn that our car is in fact broken, i.e., suppose that we add $Broken$ to our knowledge base. Somewhat surprisingly, the resulting theory turns out to be inconsistent because we still have no indication of any problem with either battery or tires and thus $\mathcal{B}\neg FlatTire$ and $\mathcal{B}\neg BadBattery$ continue to hold true.

A common-sense approach suggests that in order to avoid such inconsistencies we should refrain from adopting beliefs that contradict the existing factual information or are mutually contradictory. In this particular case, we could conclude that at least one of our initial beliefs (assumptions) that the car does not have a flat tire and does not have a bad battery must have been incorrect and thus has to be *revised* and *rejected*.

Accordingly, in this paper we first introduce the notion of a *careful static expansion*, a simple and yet powerful extension of the notion of a static expansion of belief theories, which enables us to incorporate *belief revision* into the framework of AEB. When applied to the above theory our approach results in two consistent careful static expansions. In one of them we believe that the battery is fine but possibly the tires are not, and, in the other we believe that the tires are fine but possibly the battery is dead. When taken together,

[4] In other words, if we hypothetically assume $\neg FlatTire$ and $\neg BadBattery$ the conclusion follows.

the two expansions imply that most likely either the tires or the battery, but not both, are to blame for the car's trouble. They represent therefore an intuitively appealing approach of rejecting those beliefs that contradict factual information, while keeping all the remaining ones intact. We prove that every consistent belief theory has a consistent careful static expansion. This result demonstrates that we can always assign a reasonable set of revised beliefs to any belief theory and underscores the important role played by belief revision in commonsense reasoning. We also show that every consistent static expansion of a belief theory T is also a careful static expansion of T and therefore the class of careful static expansions extends the class of consistent static expansions. Moreover, for a broad class of affirmative belief theories, defined below, careful static expansions coincide with static expansions.

Belief revision based on the notion of a careful static expansion can be applied to various reasoning domains. In this paper we illustrate its natural application to the domain of *diagnosis*. Here the fact that all consistent theories have consistent careful expansions plays a crucial role because it is imperative that we should be able to derive a reasonable set of conclusions (diagnoses) from any given knowledge base T even though the observable facts may appear to contradict beliefs resulting from default assumptions contained in T.

Careful static expansions represent a form of belief revision in which the rational epistemic agent abstains from believing formulae which, when believed, would lead to a contradiction. However, simply refraining from believing in certain formulae does not fully take into account all the consequences of the withholding such beliefs. For example, faced with the fact that the car does not run we may decide to revise our belief that the car is not broken (cf. Example 7). However, that should also compel us to refrain from believing that the car does not need to be fixed, i.e. that the car is broken should be admitted as a possibility as well. We propose a natural solution to this problem using the previously introduced notion of careful static expansion. The proposed approach is based on the appropriate *revision* of the *original theory* itself instead of just the revision of our beliefs about it. Specifically, we change the theory by adding to it new information that justifies withholding of contradictory beliefs. In other words, we compile into the theory the knowledge that prevents the same belief inconsistencies from occurring again, by allowing that the car may be broken.

Finally, we observe that in some application domains beliefs can logically depend on other beliefs, which may be viewed as more basic and sometimes considered to be non-revisable. For example, this is true when diagnosing faults in a device: causally deeper component faults are sometimes preferred over surface faults, that are simply consequences of the former. In such cases, one would like to control the level at which diagnosis is performed, by eliminating diagnoses which do not focus on the causally deeper faults. More generally, any revision of beliefs should comply with the logical dependency of beliefs. We illustrate how one can express such dependencies in AEL by means of the so called *Belief Completion Clauses*. These clauses essentially state that a revision of some beliefs requires a revision of beliefs on which they logically depend.

Because of its generality, this method of specifying the logical level of revision in belief theories can be employed to explain and justify, via the embedding of logic programs into AEB, the meta-linguistic devices used for controlling abduction, view updates and contradiction removal in logic programs. However, the lack of space precludes us from presenting these results in here. They will be presented, together with other results relating our work to other approaches to belief revision, in the forthcoming paper [APP95]. In particular, we show in there that the contradiction removal semantics for non-disjunctive extended logic programs, introduced in [PAA91, PA94], can be isomorphically embedded into the more general framework of the Autoepistemic Logic of Beliefs.

2 Autoepistemic Logic of Beliefs

We first briefly recall the definition and basic properties of the *Autoepistemic Logic of Beliefs*, AEB. The language of AEB, is a propositional modal language, $\mathcal{K}_{\mathcal{B}}$, with standard connectives ($\vee$, $\wedge$, $\supset$, $\neg$), the propositional letter $\bot$ (denoting *false*) and a modal operator $\mathcal{B}$, called the *belief* operator. The atomic formulae of the form $\mathcal{B}F$, where F is an arbitrary formula of $\mathcal{K}_{\mathcal{B}}$, are called *belief atoms*. The formulae of $\mathcal{K}_{\mathcal{B}}$ in which $\mathcal{B}$ does not occur are called *objective* and the set of all such formulae is denoted by $\mathcal{K}$. Any theory T in the language $\mathcal{K}_{\mathcal{B}}$ is called an *autoepistemic theory of beliefs*, or, briefly, a *belief theory*.

Definition 1 (Belief Theory). By an *autoepistemic theory of beliefs*, or just a *belief theory*, we mean an arbitrary theory in the language $\mathcal{K}_{\mathcal{B}}$, i.e., a (possibly infinite) set of arbitrary clauses of the form:

$$B_1 \wedge ... \wedge B_k \wedge \mathcal{B}G_1 \wedge ... \wedge \mathcal{B}G_l \wedge \neg\mathcal{B}F_1 \wedge ... \wedge \neg\mathcal{B}F_n \supset A_1 \vee ... \vee A_m$$

where $k, l, m, n \geq 0$, A_is and B_is are objective atoms and F_is and G_is are arbitrary formulae of $\mathcal{K}_{\mathcal{B}}$. Such a clause says that if the B_is are true, the G_is are believed, and the F_is are not believed then one of the A_is is true.

By an *affirmative belief theory* we mean any belief theory all of whose clauses satisfy the condition that $m > 0$. □

Observe that arbitrarily deep level of *nested beliefs* is allowed in belief theories. We assume the following two simple axiom schemata and one inference rule describing the arguably obvious properties of belief atoms:

(D) Consistency Axiom:

$$\neg\mathcal{B}\bot \tag{1}$$

(K) Normality Axiom: For any formulae F and G:

$$\mathcal{B}(F \supset G) \supset (\mathcal{B}F \supset \mathcal{B}G) \tag{2}$$

(N) Necessitation Rule: For any formula F:

$$\frac{F}{\mathcal{B}F} \tag{3}$$

The first axiom states that tautologically false formulae are *not* believed. The second axiom states that if we believe that a formula F implies a formula G and if we believe that F is true then we believe that G is true as well. The necessitation inference rule states that if a formula F has been proven to be true then F is believed to be true.

Definition 2 (Formulae Derivable from a Belief Theory). For any belief theory T, we denote by $Cn_*(T)$ the smallest set of formulae of the language $\mathcal{K}_{\mathcal{B}}$ which contains the theory T, all the (substitution instances of) the axioms (K) and (D) and is closed under standard propositional consequence and under the necessitation rule (N).

We say that a formula F is *derivable* from theory T in the logic AEB if F belongs to $Cn_*(T)$. We denote this fact by $T \vdash_* F$. We call a belief theory T *consistent* if the theory $Cn_*(T)$ is consistent. Consequently, $Cn_*(T) = \{F : T \vdash_* F\}$. Moreover, T is consistent if and only if $T \not\vdash_* \bot$. □

Remark. It is easy to see that, in the presence of the axiom (K), the axiom (D) is equivalent to the axiom:

$$\mathcal{B}F \supset \neg\mathcal{B}\neg F. \tag{4}$$

stating that if we believe in a formula F then we do *not* believe in $\neg F$.

For readers familiar with modal logics it should be clear by now that we are, in effect, considering here a *normal* modal logic with one modality $\mathcal{B}$ which satisfies the consistency axiom (D) [MT94]. The axiom (K) is called "normal" because all normal modal logics satisfy it [MT94]. □

2.1 Intended Meaning of Belief Atoms

In general, belief atoms $\mathcal{B}F$ can be given different intended meanings. In this paper, the intended meaning of belief atoms $\mathcal{B}F$ is based on Minker's *GCWA* (see [Min82, GPP89]) or McCarthy's *Predicate Circumscription* [McC80], and is described by the principle of *predicate minimization*:

$$\mathcal{B}F \;\equiv\; F \text{ is minimally entailed } \;\equiv\; F \text{ is true in all minimal models.}$$

Accordingly, beliefs considered in this paper can be called *minimal beliefs*.

We now give a precise definition of minimal models and minimal entailment. Throughout the paper we represent *models* as (consistent) *sets of literals*. An atom A is *true* in a model M if and only if A belongs to M. An atom A is *false* in a model M if and only if $\neg A$ belongs to M. A model M is *total* if for every atom A either A or $\neg A$ belongs to M. Otherwise, the model is called *partial*. Unless stated otherwise all models are assumed to be total. A (total) model M is *smaller* than a (total) model N if it contains fewer positive literals (atoms). For convenience, when describing models we usually list *only* those of their members that are *relevant* to our considerations, typically those whose predicate symbols appear in the theory that we are currently discussing.

Definition 3 (Minimal Models). [Prz94b, Prz94a] By a *minimal model* of a belief theory T we mean a model M of T with the property that there is *no* smaller model N of T which coincides with M on belief atoms $\mathcal{B}F$. If a formula F is true in all minimal models of T then we write: $T \models_{\min} F$ and say that F is *minimally entailed* by T. □

For readers familiar with *circumscription*, this means that we are considering predicate circumscription $CIRC(T; \mathcal{K})$ of the theory T in which atoms from the objective language $\mathcal{K}$ are minimized while the belief atoms $\mathcal{B}F$ are fixed:

$$T \models_{\min} F \equiv CIRC(T; \mathcal{K}) \models F.$$

In other words, minimal models are obtained by first assigning *arbitrary* truth values to the belief atoms and then *minimizing* objective atoms.

2.2 Static Autoepistemic Expansions

Like in Moore's Autoepistemic Logic, also in the Autoepistemic Logic of Beliefs we introduce sets of beliefs that an ideally rational and introspective agent may hold, given a set of premises T. We do so by defining *static autoepistemic expansions* $T^\diamond$ of T, which constitute plausible sets of such rational beliefs.

Definition 4 (Static Autoepistemic Expansion). [Prz94b, Prz94a] A belief theory $T^\diamond$ is called a *static autoepistemic expansion* of a belief theory T if it satisfies the following fixed-point equation:

$$T^\diamond = Cn_*(T \cup \{\mathcal{B}F : T^\diamond \models_{\min} F\}),$$

where F ranges over all formulae of $\mathcal{K}_\mathcal{B}$. □

The definition of static autoepistemic expansions is based on the idea of building an expansion $T^\diamond$ of a belief theory T by closing it with respect to: (i) the derivability in the logic AEB, and, (ii) the addition of belief atoms $\mathcal{B}F$ satisfying the condition that the formula F is minimally entailed by $T^\diamond$. Consequently, the definition of static expansions *enforces* the intended meaning of belief atoms described above. Note that negations $\neg\mathcal{B}F$ of the remaining belief atoms are not *explicitly* added to the expansion although some of them will be forced in by the Normality and Consistency Axioms (2) and (1).

It turns out that every belief theory T in AEB has the *least* (in the sense of set-theoretic inclusion) static expansion $T^\diamond$ which has an *iterative* definition as the *least fixed point* of the monotonic belief closure operator Ψ_T:

$$\Psi_T(S) = Cn_*(T \cup \{\mathcal{B}F : S \models_{\min} F\}),$$

where S is an arbitrary belief theory and the Fs range over all formulae of $\mathcal{K}_\mathcal{B}$.

Theorem 5 (Least Static Expansion). *[Prz94b, Prz94a] Every belief theory T in AEB has the* least *static expansion, namely, the least fixed point $T^\diamond$ of the monotonic belief closure operator Ψ_T.*

Moreover, the least static expansion $T^\diamond$ of a belief theory T can be constructed as follows. Let $T^0 = Cn_*(T)$ and suppose that T^α has already been defined for any ordinal number $\alpha < \beta$. If $\beta = \alpha + 1$ is a successor ordinal then define:

$$T^{\alpha+1} = \Psi_T(T^\alpha) = Cn_*(\ T \cup \{\mathcal{B}F : T^\alpha \ \models_{\min}\ F\}\ ),$$

where F ranges over all formulae in $\mathcal{K}_{\mathcal{B}}$. Else, if β is a limit ordinal then define $T^\beta = \bigcup_{\alpha<\beta} T^\alpha$.

The sequence $\{T^\alpha\}$ is monotonically increasing and has a unique fixed point $T^\diamond = T^\lambda = \Psi_T(T^\lambda)$, for some ordinal λ. For finite theories T the fixed point $T^\diamond$ is reached after finitely many steps. □

Observe that the *least* static autoepistemic expansion $T^\diamond$ of T contains therefore those and only those formulae which are true in *all* static autoepistemic expansions of T. It defines the so called *static semantics* of a belief theory T. It is easy to verify that a belief theory T either has a *consistent* least static expansion $T^\diamond$ or it does *not* have any consistent static expansions at all. Moreover, least static expansions of *affirmative* belief theories are always consistent [Prz94b, Prz94a].

Example 1. Consider the following belief theory T:

$$\begin{array}{l} Car \\ Car \wedge \mathcal{B}\neg Broken \supset Runs \end{array}$$

For simplicity, when describing static expansions of this and other examples we list only those elements of the expansion that are "relevant" to our discussion. In particular, we usually omit nested beliefs. In order to iteratively compute the least static expansion $T^\diamond$ of T we first let $T^0 = Cn_*(T)$. Let us observe that $T^0 \models Car$ and $T^0 \models_{min} \neg Broken$. Indeed, in order to find minimal models of T^0 we need to assign an *arbitrary* truth value to the only belief atom $\mathcal{B}\neg Broken$, and then *minimize* the objective atoms $Broken$, Car and $Runs$. We easily see that T^0 has the following two minimal models (truth values of the remaining belief atoms are irrelevant and are therefore omitted):

$$\begin{array}{l} M_1 = \{\mathcal{B}\neg Broken, Car,\ Runs,\ \neg Broken\}; \\ M_2 = \{\neg\mathcal{B}\neg Broken, Car,\ \neg Runs,\ \neg Broken\}. \end{array}$$

Since in both of them Car is true, and $Broken$ is false, we deduce that $T^0 \models_{min} Car$ and $T^0 \models_{min} \neg Broken$. Consequently, since $T^1 = \Psi_T(T^0) = Cn_*(T \cup \{\mathcal{B}F : T^0 \models_{\min} F\})$, we obtain:

$$T^1 = Cn_*(T \cup \{\mathcal{B}Car, \mathcal{B}\neg Broken\}).$$

Since $T^1 \models Runs$ and $T^2 = \Psi_T(T^1) = Cn_*(T \cup \{\mathcal{B}F : T^1 \models_{\min} F\})$, we obtain:

$$T^2 = Cn_*(T \cup \{\mathcal{B}Car, \mathcal{B}\neg Broken, \mathcal{B}Runs\}).$$

It is easy to check that $T^2 = \Psi_T(T^2)$ is a fixed point of Ψ_T and therefore $T^\diamond = T^2 = Cn_*(T \cup \{\mathcal{B}Car, \mathcal{B}\neg Broken, \mathcal{B}Runs\})$ is the least static expansion of T. The static semantics of T asserts our belief that the car is not broken and thus runs fine. One easily verifies that T does not have any other (consistent) static expansions. □

3 Belief Revision

While static expansions seem to provide a natural and intuitive semantics for many (consistent) belief theories (in particular, for all affirmative belief theories) they often lead to inconsistent expansions for theories in which (subjective) beliefs clash with the observable (objective) facts or with some other beliefs.

Example 2. Consider again the simple belief theory introduced in Example 1. As we have seen, its static semantics implies that we believe that the car is not broken and thus runs fine. Suppose, however, that upon inspection we found out that the car actually *does not* run:

$$\neg Runs.$$

It is clear that the resulting new belief theory does not have any consistent static expansions. Indeed, since there is no evidence that the car is broken, $Broken$ is false in all minimal models and thus $\mathcal{B}\neg Broken$ is derivable. This implies $Runs$ and thus results in a contradiction. In other words, our belief that the car is not broken and thus runs, based on the fact that there is no evidence to the contrary, apparently contradicts the objective fact that the car does not run.

In view of the contradictory factual information that the car does not run, we could very well conclude that our initial belief (assumption) that the car is not broken must have been incorrect and thus has to be *revised* and *rejected*. □

Example 3. Consider now the belief theory discussed in the introduction:

$$\begin{array}{l} \mathcal{B}\neg FlatTire \wedge \mathcal{B}\neg BadBattery \supset \neg Broken \\ Broken, \end{array}$$

which says that, in the absence of any indication that something is wrong with the tires or with the battery, we can safely conclude that the car is not broken, and yet the fact is that it is broken. This theory, again, does not have any consistent static expansions because both $\neg FlatTire$ and $\neg BadBattery$ are minimally entailed and thus the premise $\mathcal{B}\neg FlatTire \wedge \mathcal{B}\neg BadBattery$ is derivable. This implies $\neg Broken$ and results in a contradiction.

Again too, a natural way to remedy this problem is to conclude that, in view of the contradictory objective information that the car is broken, at least one of our initial beliefs (assumptions) that the car does not have a flat tire and does not have a bad battery must have been incorrect and thus has to be *revised* and *rejected*. □

3.1 Careful Static Autoepistemic Expansions

The approach illustrated in the previous two examples is based on the idea of *rejecting* or *revising* beliefs that contradict the existing factual information or are mutually contradictory. It leads to a simple modification of the definition of static expansions which results in a natural and potent framework for belief revision in AEL.

Definition 6 (Careful Static Autoepistemic Expansion). A belief theory $T^\diamond$ is called a *careful static autoepistemic expansion* of a belief theory T if it satisfies the following fixed-point equation:

$$T^\diamond = Cn_*(T \ \cup \{\mathcal{B}F : T^\diamond \models_{\min} F \ \text{ and } \ T^\diamond \cup \ \{\mathcal{B}F\} \text{ is consistent}\}),$$

where F ranges over all formulae of $\mathcal{K}_\mathcal{B}$. □

The only difference between the definition of static expansions and careful static expansions is the requirement that only those belief atoms $\mathcal{B}F$ should be added to the expansion whose addition does not lead to a contradiction. Recall that, by definition, $T^\diamond \cup \{\mathcal{B}F\}$ is consistent if and only if $Cn_*(T^\diamond \cup \{\mathcal{B}F\})$ is consistent.

Example 4. It is easy to see that the theory considered in Example 2 has precisely one careful static expansion, namely $T^\diamond = Cn_*(T \ \cup \{\mathcal{B}Car,\ \mathcal{B}\neg Runs\})$, which does not include any beliefs about the car being broken and corresponds therefore to the intuitive approach of rejecting beliefs that contradict existing factual information. □

Example 5. On the other hand, the theory considered in Example 3 has precisely two careful static expansions namely:

$$T_1^\diamond = Cn_*(T \ \cup \{\mathcal{B}Broken,\ \mathcal{B}\neg FlatTire\}),$$
$$T_2^\diamond = Cn_*(T \ \cup \{\mathcal{B}Broken,\ \mathcal{B}\neg BadBattery\}),$$

which reflect the fact that one of the assumptions about not having a bad battery or not having a flat tire has to be rejected while the other can be kept without causing any inconsistency. The resulting careful static semantics implies therefore $\mathcal{B}\neg FlatTire \vee \mathcal{B}\neg BadBattery$ and thus suggests that most likely the car does not have both a bad battery and a flat tire. It represents the intuitively appealing approach of rejecting only those beliefs that contradict factual information, while keeping all the remaining ones intact. □

As the previous examples demonstrate, careful static expansions no longer lead to inconsistencies when we add to our knowledge facts that seem to contradict our (default) beliefs. More generally, it turns out that every consistent belief theory has a consistent careful static expansion.

Theorem 7 (Fundamental Theorem of Belief Revision).
Every consistent belief theory has a consistent careful static expansion. □

This result demonstrates that we can always assign a reasonable set of revised beliefs to any belief theory and thus underscores the important role played by *belief revision* in commonsense reasoning. It is also of crucial importance in applications of belief revision, such as the application to *diagnosis* illustrated below, where it is imperative that we should be able to derive a reasonable set of conclusions (diagnoses) from any given knowledge base T even though the observable facts may appear to contradict beliefs resulting from default assumptions contained in T.

The class of careful static expansions extends the class of consistent static expansions.

Theorem 8. *Every consistent static expansion of a belief theory T is also a careful static expansion of T. Moreover, for affirmative belief theories, the notions of a consistent static expansion and a careful static expansion coincide.* □

3.2 Application to Diagnosis

Belief revision based on the notion of a careful static expansion can be applied to various reasoning domains. Below we illustrate its application to the domain of diagnosis.

For any careful static expansion $T^\diamond$ of a belief theory T the set $\mathcal{R}(T^\diamond) = \{F : T^\diamond \models_{min} F$ and yet $\mathcal{B}F \notin T^\diamond\}$, namely, the set of those formulae F which should be believed in (because F is minimally entailed) in the expansion $T^\diamond$, and yet are not believed in $T^\diamond$ (because of the resulting inconsistency), plays an important diagnostic role by constituting the set of *possibly false assumptions*.

Definition 9 (Revision Set of a Careful Expansion). The revision set $\mathcal{R}(T^\diamond)$ of the careful static autoepistemic expansion $T^\diamond$ of a belief theory T is defined by:

$$\mathcal{R}(T^\diamond) = \{F : T^\diamond \models_{\min} F \quad \text{and} \quad \mathcal{B}F \notin T^\diamond\}. \quad \square$$

Clearly, a careful static expansion is a (regular) static expansion if and only if its revision set is empty.

Example 6. Consider the careful static expansions of the theory discussed in Example 3:

$$T_1^\diamond = Cn_*(T \cup \{\mathcal{B}Broken,\ \mathcal{B}\neg FlatTire\})$$
$$T_2^\diamond = Cn_*(T \cup \{\mathcal{B}Broken,\ \mathcal{B}\neg BadBattery\})$$

Their revision sets are:

$$\mathcal{R}(T_1^\diamond) = \{\neg BadBattery\}$$
$$\mathcal{R}(T_2^\diamond) = \{\neg FlatTire\}$$

i.e. in $T_1^\diamond$ we refrain from believing $\neg BadBattery$, while in $T_2^\diamond$ we refrain from believing $\neg FlatTire$. As a result, the first revision set suggests that our assumption that the car does not have a bad battery may have been wrong and the second revision set suggests that our assumption that the car does not have a flat tire may have been incorrect. Both of them together provide us with a useful diagnosis of possible reasons why the car does not work. □

4 Belief Revision by Theory Change

In this section we study the issue of belief revision by theory revision, as opposed to belief revision by rejection of contradictory beliefs which was discussed in the previous section.

As remarked earlier, careful static expansions represent a form of belief revision where the rational epistemic agent abstains from believing formulae which, if believed, would lead to contradiction. However, simply refraining from believing in certain formulae is often not enough, as it does not fully take into account all the consequences of withholding such beliefs. In order to produce such consequences we must *revise* the theory by adding to it some statements that justify not holding the contradictory beliefs. In other words, we must compile into the theory *additional* knowledge that will prevent the detected belief inconsistency from occurring. This knowledge is gathered by analyzing the causes of inconsistencies.

Example 7. Suppose that to the theory of Example 2 we add:

$$Car \wedge Broken \supset FixIt$$

It is easy to check that the resulting theory T has a single careful expansion:

$$T^{\diamond} = Cn_{*}(T \cup \{\mathcal{B}Car, \mathcal{B}\neg Runs, \mathcal{B}\neg FixIt\})$$

Even though $\neg Broken$ is true in all minimal models of the expansion, $\mathcal{B}\neg Broken$ is not added since it leads to inconsistency. Since $Broken$ is no longer believed to be false, one would intuitively expect $\neg FixIt$ not to be believed either. However, this is not the case in the careful static expansion above. Indeed, the expansion reflects only the fact that the agent must refrain from believing formulae that lead to contradiction. It does not invalidate the reasons that have led to such beliefs. In our example, we believed in the car not being broken because of the lack of evidence showing otherwise. This lack of evidence must therefore be invalidated by admitting the possibility that the car might in fact be broken.

This is the stance taken by most belief revision systems, where the outcome of revision is a modified theory, in which contradiction is avoided by eliminating the reasons for contradictory beliefs. It is clear that the only way of inhibiting $\neg Broken$ from being believed in static expansions is by introducing some evidence for $Broken$ to be true. This evidence could, for example, be stated in the form that $Broken$ is in fact true. However, this appears too strong in view of the fact that we do not know for sure that the car is indeed broken: absence of belief in $\neg Broken$ does not warrant jumping to such a conclusion. An alternative is to simply state that $Broken$ is *possible*, in which case, $\neg Broken$ would no longer be minimally entailed, and we would no longer believe it. Moreover, $\neg FixIt$ would no longer be minimally entailed, and thus would no longer be believed. □

Careful expansions already identify and inhibit the addition of beliefs that lead to contradiction. It is thus an easy matter to determine which sets of formulae do lead to contradiction: they are the revision sets $\mathcal{R}(T^{\diamond})$ of careful expansions $T^{\diamond}$.

4.1 Revised Static Expansions

Given a careful expansion $T^\diamond$ of a belief theory T one can revise T by adding to it the "possibility of F being false" for every F in the revision set $\mathcal{R}(T^\diamond)$. How can this be done?

Most belief revision systems take the position that if the belief in a given formula F leads to contradiction then its complement $\neg F$ should be assumed to be true. In our opinion this is, in general, unwarranted. First of all, it is not necessary to do so in order to inhibit the belief. Moreover, it is unwarranted to jump to a conclusion that some formula is true simply because belief in its falsity would lead to contradiction. That would be tantamount to adopting on our beliefs the law of the excluded middle, i.e. imposing $\mathcal{B}F \vee \mathcal{B}\neg F$.

In Example 7, we simply would like to prevent $\neg Broken$ from being believed. Given the meaning of beliefs, this can be arranged by changing the theory just enough so that "$Broken$ is no longer false in all minimal models" or, equivalently, by "guaranteeing the existence of a minimal model in which $Broken$ is true". Technically, this is achievable by adding to the theory the clause $Broken \vee Maybe\_Not(Broken)$, where $Maybe\_Not(Broken)$ is an atom not occurring elsewhere in the theory, and thus not constrained in value. This clause can be read as *"$Broken$ is possible"*. Intuitively, this constitutes the "minimal" change of the theory ensuring that contradiction is removed. Indeed, believing $\neg Broken$ leads to contradiction and therefore $Broken$ should be possible, which effectively and declaratively prevents believing in $\neg Broken$.

For the sake of modularity, instead of adding the clauses of the form $F \vee Maybe\_Not(F)$ (where $Maybe\_Not(F)$ is an atom not occurring elsewhere in the theory), we prefer the addition of $Possible(F)$, where $Possible(F)$ is defined by[5]:

$$Possible(F) \equiv F \vee Maybe\_Not(F) \tag{5}$$

Definition 10 (Revision of a Belief Theory). A belief theory T_r is a *revision of a consistent belief theory* T if and only if

$$T_r = T \cup \{Possible(\neg F) : F \in \mathcal{R}(T^\diamond)\}$$

for some careful static expansion $T^\diamond$ of T. □

Theorem 11 (Revised Static Autoepistemic Expansion). *Let T_r be a revision of a consistent belief theory T. Then T_r is consistent and has consistent static autoepistemic expansions. We call the least static expansion of T_r, whose existence is guaranteed by Theorem 5, a* revised static autoepistemic expansion of T. □

Theorem 12 (Relation to Careful Expansions). *Let $T_r^\diamond$ be a revised static expansion of a consistent belief theory T. There exists a careful expansion $T^\diamond$ of T such that $T_r^\diamond \subseteq T^\diamond$.* □

[5] This will allow us later on [APP95] to use different, simpler, and more specific definitions of $Possible(F)$ in less general classes of theories (such as logic programs) in which there are other methods of ensuring that F is possible.

Example 8. The only revision of the theory T from Example 7 is given by $T_r = T \cup \{Possible(Broken)\}$. Accordingly, the only revised static expansion of T is:

$$T_r^\diamond = Cn_*(T \cup \{Possible(Broken)\} \cup \{\mathcal{B}Car, \mathcal{B}\neg Runs\})$$

It is easy to see that there are minimal models of the theory in which $Broken$ is true, and therefore, since Car is true in all models, those models include $FixIt$ too. Thus, neither $\mathcal{B}\neg Broken$ nor $\mathcal{B}\neg FixIt$ are added to the expansion. □

Example 9. The revisions of theory T from Example 3 are $T_{r_1} = T \cup \{Possible(BadBattery)\}$, and $T_{r_2} = T \cup \{Possible(FlatTire)\}$. Thus, the revised static expansions are:

$$T_{r_1}^\diamond = Cn_*(T \cup \{Possible(BadBattery), \mathcal{B}Broken, \mathcal{B}\neg FlatTire\})$$
$$T_{r_2}^\diamond = Cn_*(T \cup \{Possible(FlatTire), \mathcal{B}Broken, \mathcal{B}\neg BadBattery\})$$

Each of them constitutes a diagnosis of a possible problem with the car. □

4.2 Controlling the Level of Diagnosis

The belief in a formula may be conditional upon the belief in another formula. This is particularly true when diagnosing faults in a device: causally deeper component faults are sometimes preferred over less deep faults, that are simply consequences of the former. In such cases, one would like to control the level over which diagnosis is performed, by preventing diagnoses which do not focus on the causally deeper faults. We now demonstrate that revised static expansion have sufficient expressive power to control the level of diagnosis.

Example 10. The theory T:

$$\begin{array}{ll} \neg Runs & FlatTire \supset Broken \\ \mathcal{B}\neg Broken \supset Runs & BadBattery \supset Broken \end{array}$$

has a single revision: $T \cup \{Possible(Broken)\}$. The revised static expansion contains both $\mathcal{B}\neg FlatTire$ and $\mathcal{B}\neg BadBattery$. This revision can be seen as a diagnosis of the car that just states the car might be broken.

However, in this case, one would like the diagnosis to delve deeper into the car problems, and obtain one diagnosis suggesting a possible problem with a flat tire and another suggesting a possible problem with a bad battery. This is justified by the fact that our belief in the car being broken seems to depend entirely on our belief that it either has a flat tire or a bad battery. □

To obtain this more desirable result one has to somehow ensure that instead of just withholding our belief in the car not being broken we in fact also withhold our belief that the car neither has a flat tire nor a bad battery. In other words, a revision of this theory should not be initiated by revising $Broken$ but instead it should be initiated by revising $FlatTire$ or $BadBattery$ by adding either $Possible(FlatTire)$ or $Possible(BadBattery)$.

Note that, by the rule (N) and the axiom (K), the closure of T already contains:

$$\mathcal{B}FlatTire \vee \mathcal{B}BadBattery \supset \mathcal{B}Broken.$$

Thus, belief in the truth of $Broken$ is already determined by the belief in $FlatTire$ or in $BadBattery$. But we intend to express the stronger fact that belief in the falsity of $Broken$ must also be determined by the beliefs held about the latter literals. This is ensured by stating that if both $FlatTire$ and $BadBattery$ are believed false then $Broken$ must be also believed false:

$$\mathcal{B}\neg FlatTire \wedge \mathcal{B}\neg BadBattery \supset \mathcal{B}\neg Broken \tag{6}$$

Example 11. The theory T from Example 10, augmented with clause (6) now has two revised expansions:

$$\begin{aligned} T^{\diamond}_{r_1} &= Cn_*(T \ \cup \{Possible(BadBattery), \mathcal{B}\neg Runs, \mathcal{B}\neg FlatTire\}) \\ T^{\diamond}_{r_2} &= Cn_*(T \ \cup \{Possible(FlatTire), \quad \mathcal{B}\neg Runs, \mathcal{B}\neg BadBattery\}) \end{aligned}$$

each corresponding to one of the desired deeper diagnoses.
$T \cup \{Possible(Broken)\}$ is no longer a revision because it still derives $\mathcal{B}\neg Broken$, via clause (6), and thus is inconsistent. □

Note the similarities between clause (6) and Clark's completion [Cla78] of $Broken$. Clark's completion states that if both $FlatTire$ and $BadBattery$ are false then $Broken$ is false, whilst (6) refers instead to the corresponding beliefs in their falsity. For this reason we call (6) the *belief completion clause* for $Broken$. More generally:

Definition 13 (Belief Completion Clauses). Let T be an AEB theory, and let:

$$\begin{aligned} B_{1,1} \wedge \ldots \wedge B_{1,m} \wedge \mathcal{B}\neg B_{1,m+1} \wedge \ldots \wedge \mathcal{B}\neg B_{1,n} \ &\supset \ A \\ &\ldots \\ B_{k,1} \wedge \ldots \wedge B_{k,m} \wedge \mathcal{B}\neg B_{k,m+1} \wedge \ldots \wedge \mathcal{B}\neg B_{k,n} \ &\supset \ A \end{aligned}$$

be the clauses[6] for A in T, where A is an atom, each Bi, j is a literal, and $k > 0$. The *belief completion clauses for A in T*, $BelComp(A)$, are:

$$\begin{aligned} &(\mathcal{B}\neg B_{1,1} \vee \ldots \vee \mathcal{B}\neg B_{1,m} \vee \mathcal{B}B_{1,m+1} \vee \ldots \vee \mathcal{B}B_{1,n}) \\ \wedge \ldots \wedge\ &(\mathcal{B}\neg B_{k,1} \vee \ldots \vee \mathcal{B}\neg B_{k,m} \vee \mathcal{B}B_{k,m+1} \vee \ldots \vee \mathcal{B}B_{k,n}) \supset \mathcal{B}\neg A \end{aligned}$$

If there are no clauses for A in T then its belief completion is $\mathcal{B}\neg A$. □

By adding the completion rules for an atom A, we can therefore prevent revision to be initiated in $\mathcal{B}\neg A$, i.e., in order to revise the belief in $\neg A$, beliefs in other literals on which A depends must also be revised. In diagnosis, the hierarchical component structure of artifacts naturally induces dependency levels into theories modeling them. In other words, we can impose, via belief completion clauses, the desired levels of diagnosis in artifacts.

[6] By a clause for an atom A we mean one in which A occurs positively.

References

[ADP94] J. J. Alferes, C. V. Damásio, and L. M. Pereira. SLX – a top-down derivation procedure for programs with explicit negation. In M. Bruynooghe, editor, *Int. Logic Programming Symposium.* MIT Press, 1994.

[ADP95] J. J. Alferes, C. V. Damásio, and L. M. Pereira. A logic programming system for non-monotonic reasoning. *Journal of Automated Reasoning*, Special Issue on Implementation of NonMonotonic Reasoning(14):93–147, 1995.

[Alf93] J. J. Alferes. *Semantics of Logic Programs with Explicit Negation.* PhD thesis, Universidade Nova de Lisboa, 1993.

[AP92] J. J. Alferes and L. M. Pereira. On logic program semantics with two kinds of negation. In K. Apt, editor, *International Joint Conference and Symposium on Logic Programming*, pages 574–588. MIT Press, 1992.

[AP94] J. J. Alferes and L. M. Pereira. Belief, provability and logic programs. In C. MacNish et al., editors, *Logics in AI*, pages 106-121. Springer-Verlag LNAI 838, 1994. Extended version in *Journal of Applied Nonclassical Logics*, 5(1):31–50, 1995.

[APP95] J. J. Alferes, L. M. Pereira, and T. Przymusinski. Strong and explicit negation in non-monotonic reasoning and logic programming. (in preparation), 1995.

[Cla78] K. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 293–322. Plenum Press, 1978.

[GL88] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In R. Kowalski and K. Bowen, editors, *Proceedings of the Fifth Logic Programming Symposium*, pages 1070–1080, Cambridge, Mass., 1988. Association for Logic Programming, MIT Press.

[GL90] M. Gelfond and V. Lifschitz. Logic programs with classical negation. In *Proceedings of the Seventh International Logic Programming Conference, Jerusalem, Israel*, pages 579–597, Cambridge, Mass., 1990. Association for Logic Programming, MIT Press.

[GPP89] M. Gelfond, H. Przymusinska, and T. Przymusinski. On the relationship between circumscription and negation as failure. *Journal of Artificial Intelligence*, 38:75–94, 1989.

[Lif92] V. Lifschitz. Minimal belief and negation as failure. Research report, University of Texas at Austin, 1992.

[McC80] J. McCarthy. Circumscription – a form of non-monotonic reasoning. *Journal of Artificial Intelligence*, 13:27–39, 1980.

[Min82] J. Minker. On indefinite data bases and the closed world assumption. In *Proc. 6-th Conference on Automated Deduction*, pages 292–308, New York, 1982. Springer Verlag.

[Moo85] R.C. Moore. Semantic considerations on non-monotonic logic. *Journal of Artificial Intelligence*, 25:75–94, 1985.

[MT94] W. Marek and M. Truszczynski. *Non-Monotonic Logic.* Springer Verlag, 1994.

[PA92] L. M. Pereira and J. J. Alferes. Well founded semantics for logic programs with explicit negation. In B. Neumann, editor, *European Conference on Artificial Intelligence*, pages 102–106. John Wiley & Sons, 1992.

[PA94] L. M. Pereira and J. J. Alferes. Contradiction: when avoidance equal removal. Part II. In R. Dyckhoff, editor, *Extensions of Logic Programming*, number 798 in LNAI, pages 268–281. Springer–Verlag, 1994.

[PAA91] L. M. Pereira, J. J. Alferes, and J. N. Aparício. Contradiction Removal within Well Founded Semantics. In A. Nerode, W. Marek, and V. S. Subrahmanian, editors, *Logic Programming and NonMonotonic Reasoning*, pages 105–119. MIT Press, 1991.

[PAA93] L. M. Pereira, J. N. Aparício, and J. J. Alferes. Non–monotonic reasoning with logic programming. *Journal of Logic Programming. Special issue on Nonmonotonic reasoning*, 17(2, 3 & 4):227–263, 1993.

[PDA93] L. M. Pereira, C. Damásio, and J. J. Alferes. Diagnosis and debugging as contradiction removal. In L. M. Pereira and A. Nerode, editors, *2nd International Workshop on Logic Programming and NonMonotonic Reasoning*, pages 316–330. MIT Press, 1993.

[Prz90] T. C. Przymusinski. The well-founded semantics coincides with the three-valued stable semantics. *Fundamenta Informaticae*, 13(4):445–464, 1990.

[Prz91a] T. C. Przymusinski. Autoepistemic logics of closed beliefs and logic programming. In A. Nerode, W. Marek, and V.S. Subrahmanian, editors, *Proceedings of the First International Workshop on Logic Programming and Non-monotonic Reasoning, Washington, D.C., July 1991*, pages 3–20, Cambridge, Mass., 1991. MIT Press.

[Prz91b] T. C. Przymusinski. Stable semantics for disjunctive programs. *New Generation Computing Journal*, 9:401–424, 1991. (Extended abstract appeared in: Extended stable semantics for normal and disjunctive logic programs. *Proceedings of the 7-th International Logic Programming Conference, Jerusalem*, pages 459–477, 1990. MIT Press.).

[Prz94a] T. C. Przymusinski. Autoepistemic logic of knowledge and beliefs. (in preparation), University of California at Riverside, 1994.

[Prz94b] T. C. Przymusinski. A knowledge representation framework based on autoepistemic logic of minimal beliefs. In *Proceedings of the Twelfth National Conference on Artificial Intelligence, AAAI-94, Seattle, Washington, August 1994*, page (in print), Los Altos, CA, 1994. American Association for Artificial Intelligence, Morgan Kaufmann.

[Prz94c] T. C. Przymusinski. Static semantics for normal and disjunctive logic programs. *Annals of Mathematics and Artificial Intelligence*, 1994. (in print).

[Rei78] R. Reiter. On closed-world data bases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, New York, 1978.

[Rei87] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–96, 1987.

[RLM89] A. Rajasekar, J. Lobo, and J. Minker. Weak generalized closed world assumption. *Journal of Automated Reasoning*, 5:293–307, 1989.

[RT88] K. Ross and R. Topor. Inferring negative information from disjunctive databases. *Journal of Automated Reasoning*, 4:397–424, 1988.

[VGRS90] A. Van Gelder, K. A. Ross, and J. S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 1990. Preliminary abstract appeared in Seventh ACM Symposium on Principles of Database Systems, March 1988, pp. 221–230.