Resource Bounded Belief Revision

Renata Wassermann

Institute for Logic, Language and Computation
University of Amsterdam
Plantage Muidergracht, 24
1018-TV Amsterdam
The Netherlands

e-mail: renata@wins.uva.nl

Resource Bounded Belief Revision

Abstract

The AGM paradigm for belief revision provides a very elegant and powerful framework for reasoning about idealized agents. The paradigm assumes that the modeled agent is a perfect reasoner with infinite memory. In this paper we propose a framework to reason about non-ideal agents that generalizes the AGM paradigm.

We first introduce a structure to represent an agent's belief states that distinguishes different status of beliefs according to whether or not they are explicitly represented, whether they are currently active and whether they are fully accepted or provisional. Then we define a set of basic operations that change the status of beliefs and show how these operations can be used to model agents with different capacities. We also show how different operations of belief change described in the literature can be seen as special cases of our theory.

1 Introduction

The problem of belief revision, i.e., of how the beliefs of an agent should change in the presence of new information, has been recently addressed by various authors. In most approaches, the agents are idealized in that they are assumed to have perfect recall and to hold only consistent beliefs, which are furthermore assumed to be closed under logic consequence. Harman (1986) presents an analysis of belief revision for non-ideal agents which, though informal, will be used as a guideline for a more formal proposal. In this paper we proceed towards a theory of belief revision for resource-bounded agents, in which we take into consideration the effects of both limited memory and limited capacities of inference.

The classical theory of belief change, known as the AGM paradigm (Alchourrón et al., 1985), is a theory of highly idealized reasoners. An agent's belief state is modeled by theories (sets of formulas closed under logical consequence), called *belief sets*, on which three change operations are proposed: expansion, contraction and revision. *Expansion* consists in taking the union of the prior belief set with the new belief and forming the logical closure. *Contraction* consists in deleting as many beliefs as necessary from the belief set so that the result is logically closed and does not contain a given belief. *Revision* consists in adding a belief to a belief set but in such a way that the resulting belief set is consistent, meaning that some old beliefs may have to be given up. While there is a unique way to expand belief sets, contraction and revision are not uniquely defined. The AGM paradigm provides sets of rationality postulates that any operation of contraction or revision should satisfy. For an overview of the AGM paradigm, see Gärdenfors (1988), Gärdenfors and Rott (1995), Hansson (1998).

An alternative to the use of belief sets for representing belief states is to use a set of formulas not closed under logical consequence, called a *belief base*. This alternative has been extensively studied and AGM-like operations have been defined for belief bases (Fuhrmann, 1991; Hansson, 1989; Nebel, 1992). The elements of a belief base are assumed to be in a sense more basic beliefs, from which logical consequences can be derived. The use of belief bases has clear computational advantages since it allows for a more compact representation of a belief state. It also allows for more expressive power, due to the distinction between basic and merely derived beliefs.

In this paper, we propose that the agent has a kind of short-term memory in which recently computed results are stored. When a result is important, or frequently used, the agent may store it explicitly in his long-term memory. In our framework, an agent's long-term memory is represented as a belief base, a set of formulas which is not closed under logical consequence. What distinguishes our proposal from the previous ones is the fact that changes in a belief state take place initially in a short-term memory, rather than in the long-term memory. The structure of a belief state will be formally described in section 2.

It is important to note that what we are looking for is not a limited implementation of a theory for ideal reasoning, but rather a theory for reasoners with limited resources, such as humans, computers, robots. The assumption very often made that the agent's beliefs are closed under logical rules is not only a problem from the computational point of view but there is also the question of why an agent would want to know all irrelevant consequences of his beliefs. Cherniak (1986) presented a theory for "minimal agents", agents that have the minimal abilities that are required for them to be called rational. According to Cherniak, any rational agent (limited reasoners included) must satisfy the *minimal general rationality condition*: "If agent X has a particular belief-desire set, X would undertake some, but not necessarily all, of those actions that are apparently appropriate" (page 9). From this condition Cherniak derives the *minimal inference condition*: "X would make some, but not necessarily all, of the sound inferences from the belief set that are apparently appropriate" (page 10).

Harman (1986) states some principles that should be valid for any resource-bounded agent:

1. *Clutter Avoidance:* "One should not clutter one's mind with trivialities." (page 12)

2. *Recognized Implication Principle:* "One has a reason to believe P if one recognizes that P is implied by one's view." (page 18)

3. *Recognized Inconsistency Principle:* "One has a reason to avoid be-

lieving things one recognizes to be inconsistent." (page 18)

4. *Principle of Positive Undermining:* "One should stop believing P whenever one positively believes one's reasons for believing P are no good." (page 39)

5. *Principle of Conservatism:* "One is justified in continuing fully to accept something in the absence of a special reason not to." (page 46)

6. *Interest Condition:* "One is to add a new proposition P to one's beliefs only if one is interested in whether P is true (and it is otherwise reasonable for one to believe P)." (page 55)

7. *Get Back Principle:* "One should not give up a belief one can easily (and rationally) get right back." (page 58)

In the next section, we will present a formal model of belief states, where we distinguish between beliefs that are explicit and implicit, active and inactive, provisional and accepted. In section 3, we present a set of basic operations for belief change that can be applied to belief states. These operations can be combined to form more complex operations. This is illustrated in section 4, where we show how to define local change (Hansson and Wassermann, 1998) using belief states equipped with the basic operations. We will then discuss how Harman's principles can be interpreted in this framework. In section 6, we present some conclusions and point toward future work.

In the rest of this paper, we will be working with a propositional language L, closed under the usual connectives and containing a constant $\perp$ representing falsum. We use Cn to denote a Tarski-style consequence operation.

2 Belief States

In this section we present our model for belief states. We start by introducing some distinctions between different kinds of beliefs. The example below motivates the distinctions.

Consider the following situation: Mary is going out, and her mother tells her that she should take an umbrella. Besides beliefs about other subjects, she holds the belief that if she is going to be outside for a long time, then she should take the umbrella. She also beliefs that she will be outside the whole day. If her mother had not mentioned the umbrella, she would not have thought of it. Upon it being brought to her notice, she concludes she should indeed take the umbrella.

Harman proposes that some of the agent's beliefs are explicitly represented. His definition of implicit beliefs is rather vague. Beliefs that can be inferred from explicit ones belong to this category, but not all implicit beliefs need to be so derivable. Following Harman, we will assume that there are beliefs that are explicitly represented, from which others may be inferred.

Departing from him, however, we will identify the set of the agent's implicit beliefs with the set of beliefs that can be inferred from the explicit beliefs, in accordance with the agent's abilities.

Let E be the set of the agent's explicit beliefs, and I the set of his implicit beliefs. The set I is given by $I = \mathit{Inf}^*(E) = \bigcup_{n \geq 0} \mathit{Inf}^n(E)$, where *Inf* is a function that returns the set of formulas that the agent is able to infer from a given set of formulas in one step. The set I represents the set of beliefs the agent would be able to infer from E if he were given unlimited time.

We will not restrict ourselves to a particular notion of inference but rather consider an inference function *Inf* that will depend on the agent being modeled.[1] There are some properties that we would like *Inf* to satisfy. We would like inclusion to hold, i. e., that for any X, $X \subseteq \mathit{Inf(X)}$. We want *Inf* to give us the inferences the agent can make in one step. Thus we do not want *Inf* to be idempotent.

Cherniak (1986) defines a hierarchy of rationality concepts, on top of which appear ideal agents, with belief states that are deductively closed. On the lowest level of the hierarchy appear agents that are not able to perform any inference. These agents cannot be called rational. Resource bounded agents lie somewhere in the middle of the hierarchy. Cherniak claims that a resource-bounded agent would not be called rational if he tried to make all possible inferences from his beliefs, since this would exhaust his resources without being useful (this is analogous to Harman's Principle of Clutter Avoidance). Cherniak also notes that inference does not necessarily mean the same thing for all agents: not all agents accept the laws of logic and different agents have different limitations. He speaks of *feasible inferences*.

Another claim that appears in Cherniak (1986) is that only a small part of an agent's beliefs can be activated or thought of at a given time. This relies on the distinction between long-term and short-term memory.

We will call the information that is currently available for use *active beliefs*. These may be information that still has to be checked, such as recently acquired beliefs, intermediate conclusions in an argument, beliefs related to the current topic, etc. Some elements of the set of active beliefs are not yet really believed - at least not completely - they still must be checked. Every piece of information has first to become active in order to become accepted, rejected or revised. Not all of one's beliefs are active at the same time, as the size of the set of beliefs that can be active is often restricted.

Our belief states consist of two (possibly non-disjoint) sets, the set of explicit beliefs (E) and the set of active beliefs (A), plus an inference function that determines the set of implicit beliefs (I). In figure 1 we see a representation of an agent's belief state. All changes in belief states take place in the set of active beliefs, possibly affecting the set of explicit beliefs as well.[2]

At this point it may be useful to return to our small example to illustrate the difference between explicit and active beliefs. Mary's beliefs that if she is going to be outside for a long time, then she should take the umbrella and also that she will be outside the whole day are explicit beliefs. These beliefs only become active when her mother mentions the umbrella. When Mary thinks of it, she infers that she should take the umbrella. This example shows an argument against representing belief states as logically closed sets. Mary did not hold the belief that she should take the umbrella until the time at which the inference was made. It also shows that not all beliefs are active at the same time.

Any new belief, either coming from the "outside" (new input) or from the "inside" (inference), has to survive inquiry before being incorporated into the current beliefs. Since we allow for both inconsistent beliefs and agents which are not ideal reasoners, an inference may well not be sound. That is why inferences should be at first only provisionally accepted. The depth of the inquiry is determined by the agent and his interest in the subject. Harman (1986) defines some kinds of cognitive goals that usually guide inquiry: the interest in not being inconsistent, interest in the immediate environment, interest in facilitating reasoning (if the agent believes that knowing α would help him to obtain something he desires, he will be interested in α). For instance, consider the following example: if an agent hears that it is raining outside and he intends to go out, before going out with a raincoat and umbrella, he will probably first have a look through the window in order to be sure. But if he has no intention of going out, he might simply accept the information that it is raining and go on reading his newspaper. The agent behaves more skeptically with respect to propositions that have a direct implication in his intentions and plans or about information that comes from unreliable sources.

Harman distinguishes fully accepted beliefs from what he calls working hypotheses, the former being those working hypotheses that managed to survive inquiry. We will call working hypotheses *provisional beliefs*. Provisional beliefs are those active beliefs that are not yet accepted (explicit). In a sense, they are not real beliefs, as they are still under investigation, the agent has not yet decided whether to accept them or not. An interesting question is how a provisional belief can be granted membership in the set of accepted beliefs.

Back to our example: Let p stand for "Mary should take an umbrella" and q for "Mary will be outside for a long time". Before talking to her mother, Mary's explicit beliefs contain, among others, the beliefs q and $q \rightarrow p$. The implicit beliefs contain, among others, p. The set of active beliefs is empty (actually it could probably contain some remains of other reasoning, but this is not relevant for this argument). When the mother says that Mary should take an umbrella, p becomes a provisional active, but not explicit, belief. Mary does not necessarily believe everything her mother says immediately, so that she has to think about it. This is as if she were asking herself whether she should take the umbrella. The beliefs q and $q \rightarrow p$ become active, since they are relevant for deciding whether to accept p. When Mary eventually decides to accept p, this belief is made explicit and the set of active beliefs may get new elements according to new input.

A belief state β can be represented by $\langle E, \mathit{Inf}, A \rangle$, where E is the set of

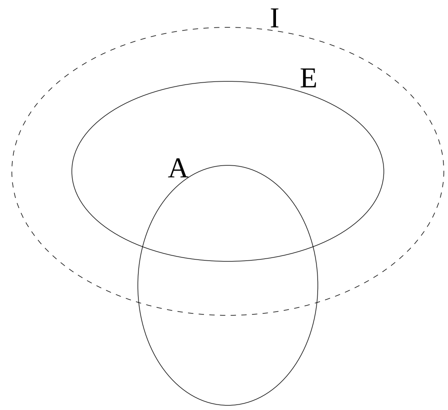

Figure 1: Structure of an agent's beliefs

the agent's explicit beliefs, *Inf* is the agents inference function and A is the set of the agent's active beliefs.[3]

As a consequence of the introduction of these distinctions between different kinds of belief, we can represent more kinds of epistemic attitudes than traditional AGM theory. In the AGM model, an agent may have one of three different epistemic attitudes concerning a sentence α (K represents the agent's belief state):

(i) α is accepted ($\alpha \in K$)

(ii) α is rejected ($\neg\alpha \in K$)

(iii) α is undetermined ($\alpha \notin K$ and $\neg\alpha \notin K$)

Our model allows for a more refined description of epistemic attitudes ($\langle E, \mathit{Inf}, A\rangle$ is the agent's belief state):

(i) α is accepted ($\alpha \in E$);

(ii) α is rejected ($\neg\alpha \in E$);

(iii) α is neither accepted nor rejected but follows from the agent's beliefs ($\alpha \in \mathit{Inf}^*(E) \setminus E$);

(iv) α is neither accepted nor rejected but can be refuted by the agent ($\neg\alpha \in \mathit{Inf}^*(E) \setminus E$);

(v) α is under consideration ($\alpha \in A \setminus E$ or $\neg\alpha \in A \setminus E$); or

(vi) none of the above, i. e., the agent is completely ignorant about α.

3 Basic Operations

In this section we define operations for changing belief states as defined in section 2.

Traditionally, revision is seen as a sequence of a contraction and an expansion (in any order). But this is not a division into simpler steps, since contraction is (computationally) as complicated as revision. We want to decompose revision and contraction in simple operations that show what happens with an agent's belief state in each step, instead of only analyzing the initial and final states.

Beliefs that are active can be forgotten or stored as explicit (but inactive) beliefs. Since the set of active beliefs is assumed to be very limited in size, there must be a mechanism that, in cases of overflow, selects which beliefs will be forgotten or stored.

The first operation we define is similar to AGM expansion in the sense that it consists in simply adding new information to a set without checking for consistency. But the operation takes the limited size of the set into account.[4] When trying to add something to a set that is already at its maximum size, some elements of the set have to be given up. This can be seen as a kind of "forgetting".

If X is a set with maximum size m and α is an element we want to add to X, then:

$$X \cup^* \{\alpha\} = X' \cup \{\alpha\}, \text{ where } X' \subseteq X,\ |X'| < m$$

Note that this operation reduces to a simple union as long as the set is not "full". Since the size m of the set is given as a parameter, the operation is more accurately denoted as $\cup^*_m$. When the set is already at its maximum size, something has to be discarded. If the set X is ordered (for example by the last time the beliefs were recalled), we can stipulate that the minimal elements of the set are the first to be dismissed, i. e., we want to ensure that if an element is dismissed, then there is no other element which is retained and that is less than the dismissed one in the order:

$$\forall y (y \in X \setminus X' \rightarrow \neg \exists x (x \in X' \wedge x < y)).$$

We define now six operations that can be applied on belief states to change the status of beliefs.

Definition 3.1 *Let $\langle E, Inf, A\rangle$ be a belief state and α a formula. We define the following operations on $\langle E, Inf, A\rangle$ (we will omit the second argument Inf since the operations defined do not affect it):*

1. *Observation ($+_o$): adds an external input to the set of active beliefs.*

   $$\langle E, A\rangle +_o \alpha = \langle E, A \cup^* \{\alpha\}\rangle$$

2. *Retrieval ($+_r$): retrieves an explicit belief into the set of active beliefs.*

   $$\langle E, A\rangle +_r \alpha = \begin{cases} \langle E, A \cup^* \{\alpha\}\rangle, & \text{if } \alpha \in E \\ \langle E, A\rangle & \text{otherwise} \end{cases}$$

3. *Acceptance ($+_a$): makes an active belief explicit.*[5]

   $$\langle E, A\rangle +_a \alpha = \begin{cases} \langle E \cup^* \{\alpha\}, A \setminus \{\alpha\}\rangle, & \text{if } \alpha \in A \\ \langle E, A\rangle & \text{otherwise} \end{cases}$$

4. *Inference* ($+_i$)*: infers something from active beliefs.*

$$\langle E, A\rangle +_i \alpha = \begin{cases} \langle E, A \cup^* \{\alpha\}\rangle, & \textit{if } \alpha \in \mathit{Inf}(A) \\ \langle E, A\rangle & \textit{otherwise} \end{cases}$$

5. *Doubting* ($+_d$)*: a belief that was accepted is questioned, becoming provisional.*

$$\langle E, A\rangle +_d \alpha = \begin{cases} \langle E \setminus \{\alpha\}, A\rangle, & \textit{if } \alpha \in A \cap E \\ \langle E, A\rangle & \textit{otherwise} \end{cases}$$

6. *Rejection* ($+_c$)*: rejects an active belief.*

$$\langle E, A\rangle +_c \alpha = \begin{cases} \langle E, A \setminus \{\alpha\}\rangle, & \textit{if } \alpha \in A \\ \langle E, A\rangle & \textit{otherwise} \end{cases}$$

The six operations defined above can be combined to model more complex operations. As an example of such a composition, consider what happens when an agent gets new information via observation. The belief will first come into the set of active beliefs through the operation $+_o$ and then the agent may accept it ($+_a$). Another example is the case of an explicit belief that becomes active (retrieval: $+_r$), when it would be expected that some implicit beliefs will also become active, i. e., the retrieval operation will be followed by an inference ($+_i$).

It is not difficult to see that, given any two belief states $\beta_1 = \langle E_1, A_1\rangle$ and $\beta_2 = \langle E_2, A_2\rangle$, there is a sequence of basic operations that takes β_1 into β_2. In fact, we can show the following:

Proposition 3.2 *The set of operations* $+_o$, $+_r$, $+_a$, $+_i$, $+_d$ *and* $+_c$ *is complete with respect to all possible changes that a belief state may undergo.*

In what follows, we will use the six operations defined also as operations that take a belief state and a finite set of formulas and return a belief state, i. e., if β is a belief state and $X = \{\chi_1, \chi_2, ..., \chi_n\}$ is a set of formulas, then $\beta \dot{+} X = \beta \dot{+} \chi_1 \dot{+} \chi_2 \dot{+} ... \dot{+} \chi_n$, where $\dot{+}$ is one of the six basic operations $+_o$, $+_r$, $+_a$, $+_i$, $+_d$ or $+_c$. Note that the operations that takes a set of formulas as input are nondeterministic, since they depend on the order in which one enumerates the elements of the set.

It is interesting to note that if we want to use the model described above to model an ideal agent, we can simulate AGM operations. This has been done in Wassermann (1997). In AGM there is no distinction between the sets of explicit and active beliefs, these sets may be infinite and the inference function used is a Tarski-style consequence operation Cn.

We will now show how the operations described in Hansson and Wassermann (1998) that make use of the set of active beliefs can be embedded in our framework.

4 Embedding Local Change

In this section, we show how to model one of the local belief change operations described in Hansson and Wassermann (1998), local contraction, in the framework defined above. All the other operations from Hansson and Wassermann (1998) can also be defined using (AGM-)expansion and local contraction.

Locally contracting a belief base B by α with respect to a set of formulas R consists in giving up enough beliefs from B such that the part of the new base that is relevant for R does not imply α. Intuitively, the set R should contain the formula α, but the formalization is general enough to allow for the use of any set of formulas. The set R should be seen as the context or topic of reasoning.

Two different constructions for local contraction are presented in Hansson and Wassermann (1998) together with sets of postulates that characterize them. We will now show how one of these constructions, namely local partial meet contraction, can be decomposed into applications of the basic operations defined above. The idea can be easily extended to the other construction (local kernel contraction) as well as to the other local operations defined in that paper.

Local operations are based on the idea of compartments. If R is a set of formulas, the *R-compartment* of a belief base B is the subset of B that is relevant for R. In Hansson and Wassermann (1998), it is assumed that a formula in B is relevant for R if this formula contributes to a proof or disproof of some element of R. This is defined using the concept of α-kernel sets which, intuitively, are the minimal subsets of a base that imply α:

Definition 4.1 *The kernel operation $\perp\!\!\!\perp$ is the operation from sets and formulas to sets of sets such that for each set of formulas B and each formula α, $X \in B \perp\!\!\!\perp \alpha$ if and only if:*

1. $X \subseteq B$
2. $\alpha \in Cn(X)$
3. *for all Y, if $Y \subset X$ then $\alpha \notin Cn(Y)$*

The elements of $B \perp\!\!\!\perp \alpha$ are called α-kernels.

The R-compartment of a belief base B is formed by taking the elements of the minimal consistent subsets of B that imply a formula of R or its negation. There is an implicit assumption here that formulas that are relevant for α are also relevant for its negation.

Definition 4.2 *Let R and B be sets of sentences. The R-compartment of B is defined as:*

$c(R, B) = \bigcup_{\alpha \in R}(\bigcup((B \perp\!\!\!\perp \alpha) \cup (B \perp\!\!\!\perp \neg\alpha) \setminus (B \perp\!\!\!\perp \bot))$

We call c a compartmentalization function.

This is only one way of defining a compartment. The representation results obtained in Hansson and Wassermann (1998) do not depend on the details of this particular construction. Another method for finding compartments of a belief base is presented in Wassermann (1998).

Local partial meet contraction is a local version of the construction for partial meet contraction given in Alchourroón et al. (1995) which also makes use of a remainder operator and a selection function. A remainder operator $\perp$ selects for every set of sentences B and every sentence α the maximal subsets of B that do not imply α. Formally:

Definition 4.3 *The remainder operation $\perp$ is the operation such that for each set of formulas B and formula α, $X \in B \perp \alpha$ if and only if:*

1. $X \subseteq B$,
2. $\alpha \notin Cn(X)$, *and*
3. $\alpha \in Cn(Y)$ *for all* Y *such that* $X \subset Y \subseteq B$.

The elements of $B \perp \alpha$ are called α-remainders.

Definition 4.4 *A selection function is a function g that selects a subset of a set of remainders such that for all sets of formulas B and formulas α,*

1. *If* $B \perp \alpha \neq \emptyset$ *then* $g(B \perp \alpha) \neq \emptyset$ *and* $g(B \perp \alpha) \subseteq B \perp \alpha$
2. *If* $B \perp \alpha = \emptyset$ *then* $g(B \perp \alpha) = B$

In an operation of partial meet contraction a selection function is used to select some of the remainders. The elements of the belief base that are not contained in all of the selected remainders are given up. In local partial meet contraction, the operation is restricted to a compartment of the belief base. If we want to contract a belief base B by the formula α with respect to a set of formulas R, the beliefs to be discarded are those in the R-compartment of B that are not contained in all the selected α-remainders of the compartment.

Definition 4.5 *We define the retain set of B given α and R as:*

$\rho_g(B, \alpha, R) = \bigcap g(c(R, B) \perp \alpha))$, *where g is a selection function.*

The discard set of B given α and R is defined by:

$\delta_g(B, \alpha, R) = c(R, B) \setminus \bigcap g(c(R, B) \perp \alpha))$, *where g is a selection function.*

Definition 4.6 [6] *Let g be a selection function. The local partial meet contraction operator with respect to R is the operator $-_R$ such that for all sets of sentences B, and sentences α:*

$B -_R \alpha = B \setminus \delta_g(B, \alpha, R)$.

The operation of local partial meet contraction leaves the irrelevant part of the belief base ($B \setminus c(R,B)$) untouched. For representation results, see Hansson and Wassermann (1998).

Let f be a function from belief bases into belief states such that for all bases B, $f(B) = \langle B, Cn, \emptyset\rangle$. We will now define an operation of local partial meet contraction on belief states that are in the image of f, i. e., belief states of the form $\langle B, Cn, \emptyset\rangle$. We will use the same symbol used for local partial meet contraction of belief bases. It will be clear from the context which of the two operations we mean.

Definition 4.7 *Let c be a compartmentalization function and g a selection function. The local partial meet contraction of a belief state $\beta = \langle B, Cn, \emptyset\rangle$ by α with respect to R is given by:*
$$\beta -_R \alpha = \beta +_r c(R,B) +_d \delta_g(B,\alpha,R) +_c \delta_g(B,\alpha,R) +_a \rho(B,\alpha,R)$$

This operation consist of retrieving the relevant compartment and deleting the beliefs contained in the discard set. The operation of doubting removes the discard set from the set of explicit beliefs, while the operation of rejecting removes the discard set from the set of active beliefs. The operation of acceptance moves the retain set into the set of explicit beliefs. Since these were already part of the set of explicit beliefs, if there is no interest in deleting these beliefs from the set of active beliefs (cf. note 3), this step may be skipped.

The operation of local partial meet contraction of belief states has the same effect on the set of explicit beliefs as the operation defined in 4.6, i. e.:

Lemma 4.8 *If R and B are sets of formulas, α is a formula and there is no maximum size for any set involved, then $f(B -_R \alpha) = f(B) -_R \alpha$*

Proof: We can see what happens to each argument of a belief state when it goes through the operation defined in 4.7. The second argument (Cn) does not change. The first argument is not affected by the retrieval ($+_r$) operation. After the doubting ($+_d$) operation, we have $B \setminus \delta_g(B,\alpha,R)$. The rejection operation does not affect the first argument and the acceptance ($+_a$) operation only adds to the first argument formulas that were already part of it. The third argument is empty before the operation. Retrieval adds $c(R,B)$ to it, doubting does not affect it, rejection deletes $\delta_g(B,\alpha,R) = c(R,B) \setminus \bigcap g(c(R,B) \perp \alpha)$ and acceptance deletes $\rho(B,\alpha,R) = \bigcap g(c(R,B) \perp \alpha)$. After the operation, the third argument is empty again.

So, if we apply a local partial meet contraction to $f(B) = \langle B, Cn, \emptyset\rangle$, we obtain $\langle B \setminus (c(R,B) \setminus D_g(B,\alpha,R), Cn, \emptyset\rangle = f(B -_R \alpha)$. □

Since all other operations defined in Hansson and Wassermann (1998) are can be obtained from applications of local contraction and expansion, we have that:

Proposition 4.9 *The theory of Local Change can be embedded in the framework of belief states with the basic operations.*

We will now show how the basic operations can be combined to form an operation of local semi-revision and apply it to our example.

The operation of semi-revising a belief base B by a sentence α consists in revising the base in a way that does not assign the highest priority to the incoming information, i. e., α may be rejected. Semi-revision consists of two phases: first the belief α is added to the base, and then the resulting base is consolidated, i. e., contracted by falsum and thus made consistent. As is the case with AGM revision, which can be defined in terms of contraction and expansion, semi-revision (?) can be defined in terms of expansion (+) and consolidation (!) via the identity Hansson (1997):

$B?\alpha = (B + \alpha)!$

The operation of local partial meet semi-revision can be defined as a composition of expansion and local partial meet consolidation (contraction by falsum):

Definition 4.10 *Let c be a compartmentalization function and g a selection function. The local partial meet semi-revision of a belief state $\beta = \langle B, Cn, \emptyset\rangle$ by α in relation to R is given by:*

$\beta ?_R \alpha = \beta +_o \alpha +_r c(R, B) +_d \delta_g(B \cup \{\alpha\}, \bot, R) +_c \delta_g(B \cup \{\alpha\}, \bot, R) +_a \rho_g(B \cup \{\alpha\}, \bot, R)$

We return to our example in order to illustrate this operation.

Suppose Mary believes that she will be outside for a long time (q), that if she stays outside for a long time, then she should take an umbrella ($q \to p$), that the moon is not made of green cheese ($\neg a$), that she loves John (b), and that Buenos Aires is the capital of Brazil (c). Her belief base is $B = \{q, q \to p, a, b, c\}$. Her belief state is given initially by: $\beta_0 = \langle B, Cn, \emptyset\rangle$. When her mother says that she should take the umbrella (p), the new belief state is given by: $\beta_1 = \beta_0 +_o p = \langle B, Cn, \{p\}\rangle$. Then the relevant beliefs are retrieved from the base: $\beta_2 = \beta_1 +_r \{q, q \to p\} = \langle B, Cn, \{p, q, q \to p\}\rangle$. Since the set of active beliefs is consistent, nothing has to be given up (note that the rest of B could still contain inconsistencies) and the result of locally consolidating gives the same belief state ($\beta_3 = \beta_2$). The active beliefs are now accepted: $\beta_4 = \beta_3 +_a \{p, q, q \to p\} = \langle B \cup \{p\}, Cn, \emptyset\rangle$.

Of course the interesting case occurs when Mary's previous beliefs are inconsistent with what her mother says. Suppose she also believed that she did not have to take an umbrella, i. e., the initial belief base was state was $B' = \{\neg p, q, q \to p, a, b, c\}$ and the initial belief state $\beta'_0 = \langle B', Cn, \emptyset\rangle$. We get $\beta'_1 = \beta'_0 +_o p = \langle B', Cn, \{p\}\rangle$ and $\beta'_2 = \beta'_1 +_r \{\neg p, q, q \to p\} = \langle B, Cn, \{p, \neg p, q, q \to p\}\rangle$. Now we have that A, the set of active beliefs, is inconsistent. For local partial meet consolidation we get: $A \perp \bot = \{\{q, q \to p, p\}, \{\neg p, q \to p\}, \{q, \neg p\}\}$. Suppose we have that $g(A \perp \bot) = \{\{q, q \to p, p\}\}$. Then the only belief given up is $\neg p$ and the new belief state is $\beta'_3 = \beta'_2 +_d \neg p +_c \neg p = \langle B' \setminus \{\neg p\}, Cn, \{q, q \to p, p\}\rangle$ and finally we have $\beta'_4 = \beta'_3 +_a \{p, q, q \to p\} = \langle (B' \setminus \{\neg p\}) \cup \{p\}, Cn, \emptyset\rangle$.

5 Harman's principles

In this section we give an interpretation for the principles in Harman (1986) that were presented in section 1 and consider how well the current proposal can be integrated with Harman's theory. Let $\langle E, Inf, A\rangle$ be a belief state.

1. *Clutter Avoidance:* This principle has as its main implication that the agent should not try to close his beliefs under logical implication, since not all consequences of the agent's explicit beliefs are useful. Clutter Avoidance does not apply to the set of implicit beliefs, that represents what the agent could (but not necessarily wants to) infer. Usually, $E \neq Inf(E)$.

2. *Recognized Implication Principle:* The agent can only recognize an implication if the premises are accepted and active. Moreover, in order to be accepted, an inference also has to be feasible, i. e., it has to be obtained by one application of *Inf*. The agent has reasons to accept a new inferred belief α if $\alpha \in Inf(A \cap E)$.

3. *Recognized Inconsistency Principle:* The agent is only aware of inconsistencies in his set of active beliefs. If an inconsistency is found, i. e., if the set of active beliefs becomes inconsistent, then there is a reason to correct it. The set $E \setminus A$ may be inconsistent, but this will not affect the reasoning.

4. *Principle of Positive Undermining:* An accepted belief can move to the set of provisional beliefs and go through inquiry again if there is evidence against it. Our theory does not say anything about what should count as evidence for or against a belief. We can imagine that a consistent set of accepted beliefs implying α could be seen as evidence for α, but there is more to evidence than this. To describe this, belief states would probably have to be enriched with a structure reflecting justifications. This is left for further work.

5. *Principle of Conservatism:* When changing his beliefs, the agent should perform only the necessary changes. Beliefs that are irrelevant for the change the agents is performing should remain untouched. Changes only take place in the set of active beliefs. The only exception is in the case where the capacity of the agent's memory is already exhausted and some beliefs have to be given up (forgotten).

6. *Interest Condition:* Here our theory does not have much to say. This principle implies that an agent's reasoning should be goal-oriented, i. e., that the agent should not make arbitrary inferences but instead pursue a goal. His interest should guide which inferences are worth making.

7. *Get Back Principle:* The agent should not give up a belief that can be reinferred from his active beliefs. This means that when giving up a

belief α, enough beliefs have to be given up so that $\alpha \notin \mathit{Inf}(A)$. But it may be the case that $\alpha \in \mathit{Inf}(E)$.

Note that the operation of local partial meet contraction (definition 4.6) agrees with our interpretation of Harman's Get Back Principle in that it satisfies local success, i. e., α is not implied by the relevant part of the resulting belief base ($\alpha \notin Cn(c(R, B -_R \alpha))$). It maybe the case that α is a consequence of the whole resulting base, i. e., that $\alpha \in Cn(B -_R \alpha)$.

6 Conclusions and Further Work

In this paper we have analyzed Harman's informal proposal for belief revision for non-ideal agents and provided a formalization that satisfies most of the principles he proposes.

We have defined a structure for belief states and a set of operations that describe how belief states can change. We have shown that these operations are sufficient to describe any change that can occur in the structure.

Our theory extends the AGM theory in the sense that it allows us to concentrate on particular subsets of an agent's beliefs and classifies them according to their status - whether they are explicitly represented, currently active and fully accepted. We have shown how AGM operations for belief change and local change as defined in Hansson and Wassermann (1998) can be seen as particular cases of our theory.

The theory also allows us to think of more general forms of belief changes by defining simple operations that work as building blocks to form more complex operations. It would be interesting to define operations that are more adequate for resource bounded agents than the ones defined in section 4. Even though those operations use the set of active beliefs in order to perform changes affecting only the relevant part of the beliefs, they rely on the classical consequence operator. Getting further away from the case of idealized agents would mean using some kind of non-classical inference operation.

As mentioned in section 5, more structure has to be added to the belief states if we want to formalize the principle of Positive Undermining. Instead of having only sets of formulas, the belief state should be organized in a way that reflects explanatory links. This too is left for further work.

Acknowledgments: I would like to thank Chris Albert, Maurice Pagnucco, Hans Rott and Frans Voorbraak for valuable comments on different earlier versions of this paper. Chris Albert and Walter Dean made my English readable. This work is supported by a grant from the Brazilian funding agency CAPES and the Dutch Research Council NWO through the PIONIER-project Reasoning with Uncertainty.

Notes

[1]The agent may be allowed to learn or revise his inference function, but we will not deal with this complication in this paper.

[2]How do beliefs become active? Two solutions for this question are presented in Hansson and Wassermann (1998) and Wassermann (1998), where two different methods for activating the relevant beliefs are described.

[3]In Wassermann (1997) yet another distinction is introduced in the belief states, between sentences that the agent is aware of believing and those he is not. Since this distinction is not needed for the results in the present paper, we adopt here a simpler definition.

[4]When we talk about the size of a set of formulas, we mean something like its complexity. The sets $\{p, q\}$ and $\{p \wedge q\}$ should have the same size. We could, for example, count the occurrence of atoms.

[5]Acceptance could also be defined without deleting the accepted belief from A, which seems to be more intuitive for human agents. The choice made here reflects our interest in artificial agents.

[6]This definition is different from the one in Hansson and Wassermann (1998), but the two definitions can be shown to be equivalent.

References

Alchourrón, C., P. Gärdenfors, and D. Makinson: 1985, On the logic of theory change, *Journal of Symbolic Logic*, 50:510–530.

Cherniak, C.: 1986, *Minimal Rationality*, MIT Press.

Fuhrmann, A.: 1991, Theory contraction through base contraction, *Journal of Philosophical Logic*, 20:175–203.

Gärdenfors, P.: 1988, *Knowledge in Flux - Modeling the Dynamics of Epistemic States*, MIT Press.

Gärdenfors, P. and H. Rott: 1995, Belief revision, In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume IV, chapter 4.2.

Hansson, S. O.: 1989, New operators for theory change, *Theoria*, 55:114–132.

Hansson, S. O.: 1997, Semi-revision, *Journal of Applied Non-Classical Logic*, 7(1-2):151–175.

Hansson, S. O.: 1998, *A Textbook of Belief Dynamics*, Kluwer Academic Press.

Hansson, S. O. and R. Wassermann: 1998, Local change, In preparation (a preliminary version appeared in the Fourth Symposium on Logical Formalizations of Commonsense Reasoning, London, 1998).

Harman, G.: 1986, *Change in View - Principles of Reasoning*, MIT Press.

Nebel, B.: 1992, Syntax based approaches to belief revision, In Gärdenfors, P., editor, *Belief Revision*. Cambridge University Press.

Wassermann, R.: 1997, Towards a theory of resource-bounded belief revision, In Engelfriet, J. and Spaan, T., editors, *Proceedings of the Accolade*.

Wassermann, R.: 1998, On structured belief bases, In *Seventh International Workshop on Nonmonotonic Reasoning*, Trento, 1998.