An information-theoretic semantics for belief change

T.A. Meyer
Computer Science Department, School of Information Technology
University of Pretoria, Pretoria, 0002, South Africa
e-mail: tmeyer@cs.up.ac.za

Abstract

Belief change is an interdisciplinary topic and is researched in departments of Computer Science, Artificial Intelligence, Philosophy, Mathematics and Engineering. It is therefore not surprising that there are a variety of approaches to defining belief change operations. In this paper we present an information-theoretic semantics for belief change. We argue that such a semantics has important implications when justifying both existing and newly constructed belief change operations.

Keywords: Artificial intelligence, knowledge representation, belief revision, theory change, belief contraction

Computing review categories: F.4.1, I.2.4

1 Introduction

Belief change is concerned with the *rational* ways in which an intelligent agent adjusts its current set of beliefs when confronted with new and possibly conflicting information. Central to the analysis of belief change is the notion of an *epistemic state*. An epistemic state contains, in one form or another, the knowledge and beliefs of an agent, together with the information needed for coherent reasoning. This includes, in particular, the strategies for performing belief change.

The enterprise of belief change was given a major thrust forward in the beginning of the eighties with a proposal known as the AGM approach to belief change [1], which is largely responsible for moving belief change from a research topic in philosophical logic to areas such as nonmonotonic reasoning [6]. With research being conducted in a number of different disciplines, it is not surprising that there are a wide variety of methods for constructing belief change operations. It is our goal in this paper to show that these different construction methods can be viewed as having an information-theoretic basis. This places us in a position to define the epistemic state of an agent information-theoretically and has a number of implications, ranging from purely philosophical issues to representational concerns.

The structure of this paper is as follows. In section 2 we describe the standard model-theoretic semantic construction of AGM belief contraction. This is followed, in section 3, by a precise description of an information-theoretic semantics and a recasting of the model-theoretic semantic methods for AGM belief contraction into such a framework. Section 4 is concerned with *epistemic entrenchment* orderings on sentences from which belief contraction operations can be obtained. While such orderings are frequently regarded as appropriate representations of the epistemic state of an agent [4], we show that these orderings are more appropriately viewed as being derived from our information-theoretic semantics for belief change. Section 5 deals with belief change operations that have come to be known as *withdrawals*, and provides a good of example of the benefits of approaching belief change from an information-theoretic point of view. Finally, section 6 shows how an information-theoretic semantics can be combined with a particular method of belief representation to yield useful belief change operations.

1.1 Preliminaries

For the rest of this paper L denotes a propositional language generated by a finite number of atoms and closed under the usual propositional connectives, $\vee$ (inclusive or), $\wedge$ (and), $\neg$ (not), $\rightarrow$ (if...then) and $\leftrightarrow$ (if and only if). A *valuation semantics* for L is provided by a set of valuations; functions from the set of atoms of L to the set $\{T, F\}$, where T denotes truth and F denotes falsity. The truth values assigned to the atoms of L by a valuation v are used to obtain a truth value of either F or T for every sentence of L under v. We say that a sentence is either true or false in a valuation v, depending on its truth value. The set of valuations of L is denoted by V. For $A \subseteq L$, we denote the *models* of A (the valuations in which the sentences in A are all true) by $M(A)$. A sentence α is said to be *semantically entailed* by a set of sentences A iff $M(A) \subseteq M(\alpha)$ and it is written as $A \vDash \alpha$. Closure under entailment is denoted by Cn. A *theory* or a *belief set* is a set $K \subseteq L$ closed under entailment. For $W \subseteq V$, we let $Th(W)$ denote the set of sentences that are true in every valuation in W. It is easily established that $Th(W)$ is a theory.

2 AGM belief contraction

AGM belief change represents an agent's beliefs as a belief set and deals with two kinds of changes: the removal of an existing belief, known as AGM contraction, and the addition of a belief that might be inconsistent with the current belief set of the agent, known as AGM revision [4]. Since belief contraction is usually regarded as more basic than revision [12], we shall only be concerned with the former. One of the guidelines of AGM belief change is the principle of Minimal Change. The intuition is that changes to an agent's current belief set should be kept to a mininum. AGM contraction can be described by the following set of postulates:

(K−1) $K - \alpha = Cn(K - \alpha)$

(K−2) $K - \alpha \subseteq K$

(K−3) If $\alpha \notin K$ then $K - \alpha = K$

(K−4) If $\nvDash \alpha$ then $\alpha \notin K - \alpha$

(K−5) If $\alpha \equiv \beta$ then $K - \alpha = K - \beta$

(K−6) If $\alpha \in K$ then $(K - \alpha) + \alpha = K$

(K−7) $(K - \alpha) \cap (K - \beta) \subseteq K - (\alpha \wedge \beta)$

(K−8) If $\beta \notin K - (\alpha \wedge \beta)$ then $K - (\alpha \wedge \beta) \subseteq K - \beta$

There are a number of different methods for constructing AGM contraction operations. We briefly describe a model-theoretic semantic method which uses total preorders on V.

From results by Grove [7] and Katsuno and Mendelzon [11], AGM contraction can be characterised semantically by a set of total preorders (i.e. connected, reflexive, transitive relations) on V. For any total preorder $\preceq$ on V, we say that $x \in W \subseteq V$ is $\preceq$-*minimal* in W iff for every $y \in W$, $x \preceq y$, and we denote the set of $\preceq$-minimal elements of $M(\alpha)$ by $M_{\preceq}(\alpha)$. For $A \subseteq L$, $\preceq$ is A-*faithful* iff $x \prec y$ for every $x \in M(A)$ and $y \notin M(A)$, and $x \preceq y$ for every $x, y \in M(A)$. Intuitively, $\preceq$ is a plausibility ordering on valuations, with a valuation lower down in the ordering being regarded as more plausible. The characterisation of AGM contraction is expressed in terms of the following identity:

(Def − from $M_{\preceq}$) $K - \alpha = Th(M(K) \cup M_{\preceq}(\neg\alpha))$

Theorem 1 *Every K-faithful total preorder defines an AGM contraction using (Def − from $M_{\preceq}$). Conversely, every AGM contraction can be defined in terms of a K-faithful total preorder using (Def − from $M_{\preceq}$).*

3 An infatom semantics for belief change

The model-theoretic semantics for AGM contraction described in section 2 is frequently viewed as a suitable representation of the epistemic states of an agent [11], [3]. But if we think of the elements of an epistemic state as objects from which (linguistic) beliefs are built up, valuations do not seem to be appropriate basic building blocks. For it is difficult to see how a valuation can be considered as a basic part of a belief expressed as a sentence in L. It is with this objection in mind that we propose the use of *infatoms* as the basic units of an epistemic state. Intuitively, infatoms are the basic independent pieces of information from which the beliefs of an agent (expressed as sentences of L) are built up. In this view, the information contained in any belief, and any belief set, is a set of infatoms. More infatoms correspond to a set of beliefs that contains more information and is logically stronger. Infatoms are *independent* of each other in the sense that it is only the set of *all* infatoms that contains too much information, leading an agent to include all sentences in its set of beliefs. Any strict subset of the set of all infatoms corresponds to a satisfiable set of beliefs. It is our contention that, since the notion of an epistemic state is so central to the study of belief change, it seems more appropriate to use a semantics based on infatoms when constructing belief change operations.

Infatoms are semantic versions of the *content elements* of Carnap and Bar-Hillel [2]. Formally, an infatom is a function i from the set of atoms of L to the set $\{I, E\}$. An infatom i sends a propositional atom α to I if i is *Included* in the information contained in α, and i sends α to E if i is *Excluded* from the information contained in α. Given a set Inf of infatoms, the *content relation* $\Vdash$ from Inf to L is then defined recursively as follows:

1. if α is an atom, then $i \Vdash \alpha$ iff $i(\alpha) = I$,
2. if $\alpha = \neg\beta$ then $i \Vdash \alpha$ iff $i \nVdash \beta$,
3. if $\alpha = \beta \vee \gamma$ then $i \Vdash \alpha$ iff $i \Vdash \beta$ and $i \Vdash \gamma$,
4. if $\alpha = \beta \wedge \gamma$ then $i \Vdash \alpha$ iff either $i \Vdash \beta$ or $i \Vdash \gamma$, or both,
5. if $\alpha = \beta \rightarrow \gamma$ then $i \Vdash \alpha$ iff $i \nVdash \beta$ and $i \Vdash \gamma$, and
6. if $\alpha = \beta \leftrightarrow \gamma$ then $i \Vdash \alpha$ iff either both $i \nVdash \beta$ and $i \Vdash \gamma$, or both $i \Vdash \beta$ and $i \nVdash \gamma$.

The *semantic content* of a set of sentences A, denoted by $C(A)$, is defined as

$$C(A) = \{i \in \text{Inf} \mid \exists\alpha \in A \text{ such that } i \Vdash \alpha\}.$$

So the semantic content of A consists of all the infatoms that are part of the information contained in at least one of the sentences in A. We shall refer to such infatoms as the *content bits* of A. We say that a set Inf of infatoms provides an *infatom semantics* for L. The *theory generated* by a set of infatoms $I \subseteq \text{Inf}$ is defined as $Th(I) = \{\alpha \mid C(\alpha) \subseteq I\}$. That is, $Th(I)$ contains all the sentences whose contents bits are included in I. Our first result about infatoms is given without proof.

Proposition 1 *Let Inf be a set of infatoms. Then* $C(\alpha \wedge \beta) = C(\alpha) \cup C(\beta)$ *and* $C(\alpha \vee \beta) = C(\alpha) \cap C(\beta)$.

It turns out that there is a natural way to associate an infatom semantics with a valuation semantics and vice versa. Given a set of valuations V, the *associated* set of infatoms is defined as $\text{Inf} = \{i_v \mid v \in V\}$, where for every $v \in V$, the *associated* infatom i_v is defined as follows: for every atom α in L, $i_v(\alpha) = I$ iff $v \notin M(\alpha)$. Similarly, given set of infatoms Inf, the *associated* set of valuations is defined as $V = \{v_i \mid i \in \text{Inf}\}$, where for every $i \in \text{Inf}$, the *associated* valuation v_i is defined as follows: for every atom α in L, $v_i(\alpha) = T$ iff $i \notin C(\alpha)$. These definitions are justified by propositions 2 and 3 below. They, in turn, rely heavily on the following lemma.

Lemma 1
1. *Let V be a set of valuations and let Inf be the associated set of infatoms. For every $\alpha \in L$ and every $v \in V$, $i_v \in C(\alpha)$ iff $v \notin M(\alpha)$.*
2. *Let Inf be a set of infatoms and let V be the associated set of valuations. For every $\alpha \in L$ and every $i \in \text{Inf}$, $v_i \in M(\alpha)$ iff $i \notin C(\alpha)$.*

The proofs follow by induction on the structure of the sentences of L.

Lemma 1 shows that an infatom is a content bit of a sentence α iff its associated valuation is not a model of α, and similarly, that a valuation is a model of α iff its associated infatom is not a content bit of α.

Proposition 2 *Let V be a set of valuations and let Inf be the associated set of infatoms.*

1. $A \vDash \beta$ *iff* $M(A) \subseteq M(\beta)$ *iff* $C(A) \supseteq C(\beta)$.
2. $Th(C(A)) = Th(M(A))$.
3. $\vDash \alpha$ *iff* $M(\alpha) = V$ *iff* $C(\alpha) = \emptyset$.
4. $C(A) = \text{Inf} \setminus \{i_u \mid u \in M(A)\}$.
5. *If* $W \subseteq V$ *and* $I = \{i_w \mid w \in W\}$ *then* $Th(W) = Th(\text{Inf} \setminus I)$.
6. $Th(M(A) \cup \{v\}) = Th(C(A) \setminus \{i_v\})$.
7. *If* $v \in V$ *then* $Th(M(A) \setminus \{v\}) = Th(C(A) \cup \{i_v\})$.

Proof
1. Suppose that $M(A) \subseteq M(\beta)$ and pick any $i_v \in C(\beta)$. Now assume that $i_v \notin C(A)$. That is, for every $\alpha \in A$, $i_v \notin C(\alpha)$. Then, by lemma 1, $v \in M(\alpha)$ for every $\alpha \in A$, and therefore $v \in M(A)$. But by supposition, $v \in M(\beta)$, and by lemma 1, $i_v \notin C(\beta)$; a contradiction. Conversely, suppose that $C(\beta) \subseteq C(A)$ and pick any $v \in M(A)$. Now assume that $v \notin M(\beta)$. By lemma 1, $i_v \in C(\beta)$. So $i_v \in C(A)$, and there is thus an $\alpha \in A$ such that $i_v \in C(\alpha)$. But by lemma 1, $v \notin M(\alpha)$, contradicting the supposition that $v \in M(A)$.
2. $\alpha \in Th(C(A))$ iff $C(\alpha) \subseteq C(A)$ iff $M(A) \subseteq M(\alpha)$ (by part (1) above) iff $\alpha \in Th(M(A))$.
3. It follows easily from the definitions of $M(\alpha)$ and $C(\alpha)$ that $M(\alpha) = V$ iff $\vDash \alpha$ and that $C(\alpha) = \emptyset$ iff $\vDash \alpha$.
4. Pick an $i_v \in C(A)$. So, there is an $\alpha \in A$ such that $i_v \in C(\alpha)$. By lemma 1, $v \notin M(\alpha)$. So $v \notin M(A)$, and thus $i_v \in \text{Inf} \setminus \{i_w \mid w \in M(A)\}$. Conversely, suppose that $i_v \in \text{Inf} \setminus \{i_w \mid w \in M(A)\}$. Then $v \notin M(A)$, and there is thus an $\alpha \in A$ such that $v \notin M(\alpha)$. Hence, by lemma 1, $i_v \in C(\alpha)$. Therefore $i \in C(A)$.
5. Suppose that $W \subseteq V$ and $I = \{i_w \mid w \in W\}$, and pick an $\alpha \in Th(W)$. So $W \subseteq M(\alpha)$. Now assume that $\alpha \notin Th(\text{Inf} \setminus I)$. That is, $C(\alpha) \nsubseteq \text{Inf} \setminus I$. In other words, there is an $i_v \in C(\alpha)$ such that $i_v \in I$. So $v \in W$ and by lemma 1, $v \notin M(\alpha)$, contradicting the fact that $W \subseteq$

$M(\alpha)$. Conversely, suppose that $\alpha \in Th(\text{Inf} \setminus I)$. So $C(\alpha) \subseteq \text{Inf} \setminus I$. Now assume that $\alpha \notin Th(W)$. Then $W \not\subseteq M(\alpha)$, and there is thus a $w \in W$ such that $w \notin M(\alpha)$. So $i_w \in I$, and by lemma 1, $i_w \in C(\alpha)$, thus contradicting the fact that $C(\alpha) \subseteq \text{Inf} \setminus I$.

6. Pick any $\beta \in Th(M(A) \cup \{v\})$. That is, $M(A) \cup \{v\} \subseteq M(\beta)$. It suffices to show that $C(\beta) \subseteq C(A) \setminus \{i_v\}$. So pick any $i_w \in C(\beta)$. By lemma 1, $w \notin M(\beta)$, and this means that $w \notin M(A) \cup \{v\}$. So $v \neq w$ (and hence $i_v \neq i_w$) and there is an $\alpha \in A$ such that $w \notin M(\alpha)$. By lemma 1 it then follows that $i_w \in C(\alpha)$, and thus that $i_w \in C(A)$. The required result then follows from the fact that $i_v \neq i_w$. Conversely, pick any $\beta \in Th(C(A) \setminus \{i_v\})$. That is $C(\beta) \subseteq C(A) \setminus \{i_v\}$. It suffices to show that $M(A) \cup \{v\} \subseteq M(\beta)$. So pick any $w \in M(A) \cup \{v\}$. Then either $w = v$ (and hence $i_v = i_w$), or $w \in M(\alpha)$ for every $\alpha \in A$. In the former case $i_w \notin C(\beta)$, and in the latter case, it follows by lemma 1 that $i_w \notin C(\alpha)$ for every $\alpha \in A$, and thus that $i_w \notin C(A)$. So either way, $i_w \notin C(\beta)$, and hence, by lemma 1, $w \in M(\beta)$.

7. Pick any $\beta \in Th(M(A) \setminus \{v\})$. That is, $M(A) \setminus \{v\} \subseteq M(\beta)$. It suffices to show that $C(\beta) \subseteq C(A) \cup \{i_v\}$. So pick any $i_w \in C(\beta)$. By lemma 1, $w \notin M(\beta)$. And this means that $w \notin M(A) \setminus \{v\}$. So there is an $\alpha \in A$ such that $w \notin M(\alpha)$ and it follows by lemma 1 that $i_w \in C(\alpha)$. But then $i_w \in C(A)$, which means we are done. Conversely, pick any $\beta \in Th(C(A) \cup \{i_v\})$. That is, $C(\beta) \subseteq C(A) \cup \{i_v\}$. It suffices to show that $M(A) \setminus \{v\} \subseteq M(\beta)$. So pick any $w \in M(A) \setminus \{v\}$. Then $w \in M(\alpha)$ for every $\alpha \in A$. By lemma 1, $i_w \notin C(\alpha)$ for every $\alpha \in A$. Therefore $i_w \notin C(A)$. Furthermore $i_w \neq i_v$, for if $i_w = i_v$, it would have meant that $w = v$, contradicting the fact that $w \in M(A) \setminus \{v\}$. And thus $i_w \notin C(\beta)$, which means, by lemma 1, that $w \in M(\beta)$.

□

Of particular interest in the proposition above are the last two parts. Part (6) shows that adding a valuation to the models of a set of sentences A is the same as removing its associated infatom from the semantic content of A. Part (7) shows that the removal, from the models of A, of a valuation v is the same as adding v's associated infatom to the semantic content of A.

The next proposition draws more parallels between valuations and infatoms.

Proposition 3 *Let Inf be a set of infatoms and let V be the associated set of valuations.*

1. $M(A) = V \setminus \{v_i \mid i \in C(A)\}$.
2. *If* $I \subseteq \text{Inf}$ *and* $W = \{w_i \mid i \in \text{I}\}$ *then* $Th(I) = Th(V \setminus W)$.
3. $Th(C(A) \cup \{i\}) = Th(M(A) \setminus \{v_i\})$.
4. $Th(C(A) \setminus \{i\}) = Th(M(A) \cup \{v_i\})$.

Proof 1. $v_i \in M(A)$ iff $v_i \in M(\alpha)$ for every $\alpha \in A$, iff $i \notin C(\alpha)$ for every $\alpha \in A$ (by lemma 1), iff $i \notin C(A)$, iff $v_i \in V \setminus \{w_i \mid i \in C(A)\}$.

2. Suppose that $I \subseteq \text{Inf}$ and $W = \{w_i \mid i \in I\}$. Now pick any $\alpha \in Th(I)$. That is, $C(\alpha) \subseteq I$. It suffices to show that $V \setminus W \subseteq M(\alpha)$. So pick any $w_i \in V \setminus W$. By the definition of W it follows that $i \notin I$. But this means that $i \notin C(\alpha)$, and by lemma 1, that $w_i \in M(\alpha)$. Conversely, pick any $\alpha \in Th(V \setminus W)$. That is, $V \setminus W \subseteq M(\alpha)$. It suffices to show that $C(\alpha) \subseteq I$. So pick an $i \notin I$. By the definition of W it follows that $w_i \in V \setminus W$. But then $w_i \in M(\alpha)$, and by lemma 1, $i \notin C(\alpha)$.

3. The proof is very similar to the proof of part (7) of proposition 2 and is omitted.
4. The proof is very similar to the proof of part (6) of proposition 2 and is omitted.

□

With the aid of these results we can now provide an infatom semantics for AGM contraction. Given a set of sentences A, we refer to a total preorder $\preceq$ on *infatoms* as *A-faithful* iff for every $i, j \notin C(A)$, $i \preceq j$ and for every $i \notin C(A)$ and $j \in C(A)$, $i \prec j$. Intuitively, such an ordering reflects the degree to which a content bit of A is entrenched in A; the higher up in the total preorder, the more entrenched it is. And as a limiting case, the infatoms that are not content bits of A are seen as the least entrenched. We denote the set of $\preceq$-minimal elements of $C(\alpha)$ by $C_{\preceq}(\alpha)$. For a belief set K, let us now define a belief removal operation in terms of a K-faithful total preorder $\preceq$ on infatoms as follows:

(Def − from $C_{\preceq}$) $K - \alpha = Th(C(K) \setminus C_{\preceq}(\alpha))$

In other words, to remove the sentence α from the belief set K, we remove the least entrenched content bits of α from K.

It is easily established that by replacing the infatoms in such a K-faithful total preorder with their associated valuations, we obtain a K-faithful total preorder on valuations. By repeated applications of part (6) of proposition 2 and part (4) of proposition 3, it then also follows that (Def − from $C_{\preceq}$) is just the infatom version of (Def − from $M_{\preceq}$). By theorem 1, the belief removal operations defined in terms of faithful total preorders on infatoms using (Def − from $C_{\preceq}$) are thus precisely the AGM contraction operations.

Although technically similar to a valuation semantics, the use of an infatom semantics seems more appropriate from an epistemological point of view, mainly because the basic building blocks are bits of information instead of valuations.

4 Epistemic entrenchment

The basic idea behind epistemic entrenchment is that some of our beliefs are more firmly entrenched than others, and we would thus be more willing to give up the latter sentences than the former if we are forced to choose. Gärdenfors and Makinson [5] and Gärdenfors [4] have defined an *epistemic entrenchment* ordering (or an EE-ordering), relative to a given belief set K, as a binary relation on L satisfying the following set of postulates (with sentences higher up in the ordering being more entrenched):

(EE1) $\sqsubseteq_{EE}$ is transitive

(EE2) If $\alpha \vDash \beta$ then $\alpha \sqsubseteq_{EE} \beta$

(EE3) $\alpha \sqsubseteq_{EE} \alpha \wedge \beta$ or $\beta \sqsubseteq_{EE} \alpha \wedge \beta$

(EE4) If $K \neq Cn(\bot)$ then $\alpha \notin K$ iff $\alpha \sqsubseteq_{EE} \beta$ for all β

(EE5) If $\alpha \sqsubseteq_{EE} \beta$ for all α then $\vDash \beta$

From results in [5], AGM contraction and epistemic entrenchment are interdefinable by means of the following two identities:

(Def − from $\sqsubseteq_{EE}$) $K - \alpha = K$ if $\alpha \notin K$ or $\vDash \alpha$, and $K - \alpha = K \cap \{\beta \mid \alpha \sqsubset_{EE} \alpha \vee \beta\}$ otherwise

(Def $\sqsubseteq_{EE}$ from −) $\alpha \sqsubseteq_{EE} \beta$ iff $\alpha \notin K - (\alpha \wedge \beta)$ or $\vDash \alpha \wedge \beta$

Theorem 2 *A removal is an AGM contraction iff it is defined in terms of an EE-ordering using (Def − from $\sqsubseteq_{EE}$). A binary relation on L is an EE-ordering iff it is defined in terms of an AGM contraction using (Def $\sqsubseteq_{EE}$ from −).*

It turns out that the EE-orderings can be defined in terms of the faithful total preorders on valuations using the following identity [4].

(Def $\sqsubseteq_{EE}$ from $M_{\preceq}$) $\alpha \sqsubseteq_{EE} \beta$ iff $\forall y \in M(\neg\beta)$ $\exists x \in M(\neg\alpha)$ such that $x \preceq y$

Theorem 3 *Every faithful total preorder defines an EE-ordering using (Def $\sqsubseteq_{EE}$ from $M_{\preceq}$). Conversely, every EE-ordering can be defined in terms of a faithful total preorder using (Def $\sqsubseteq_{EE}$ from $M_{\preceq}$).*

Viewed model-theoretically, this seems like little more than a mildly interesting technical result. With the help of lemma 1, however, we obtain the following infatom version of (Def $\sqsubseteq_{EE}$ from $M_{\preceq}$):

(Def $\sqsubseteq_{EE}$ from $C_{\preceq}$) $\alpha \sqsubseteq_{EE} \beta$ iff $\forall j \in C(\beta)$ $\exists i \in C(\alpha)$ such that $i \preceq j$

(Def $\sqsubseteq_{EE}$ from $C_{\preceq}$) states that α is at most as entrenched as β iff for every content bit of β we can find a content bit of α that is as most as entrenched as the content bit of β. In other words, the level of entrenchment of a sentence is completely determined by its least entrenched content bits. And from lemma 1 and theorem 3 it is easily established that the orderings on sentences obtained in terms of (Def $\sqsubseteq_{EE}$ from $C_{\preceq}$) using the faithful total preorders on infatoms are precisely the EE-orderings of Gärdenfors and Makinson.

5 Withdrawal

While AGM belief change has had a big influence on research involving belief change, the particular manner in which the principle of Minimal Change has been applied in the case of AGM contraction has been called into doubt. The criticism boils down to drawing into question the validity of the postulate (K$-$6), which is known as the Recovery postulate [10]. We refer to removal operations which satisfy all the AGM contraction postulates, with the possible exception of the Recovery postulate, as *principled withdrawal* operations. For long, it was regarded as quite difficult to define principled withdrawal operations which do not satisfy Recovery. This has now changed with the introduction of Rott and Pagnucco's [17] severe withdrawal, the systematic withdrawal of Meyer et al. [15], and Levi's [13] mild contraction. We focus on the first of these three.

Severe withdrawal is defined in terms of faithful total preorders on valuations. Given a total preorder $\preceq$ on valuations, define the *model-theoretic downset* $\nabla^M_{\preceq}(\alpha)$ of a sentence α as follows: $\nabla^M_{\preceq}(\alpha) = \{x \mid \exists y \in M_{\preceq}(\alpha)$, such that $x \preceq y\}$. A *severe withdrawal* is then defined in terms of a faithful total preorder $\preceq$ on valuations using the following identity:

(Def $\sim$ from $\nabla^M_{\preceq}$) $K \sim \alpha = Th(M(K) \cup \nabla^M_{\preceq}(\neg\alpha))$

As is the case with AGM contraction and epistemic entrenchment, a recasting of these definitions in terms of infatoms sheds more light on the issue. Given a total preorder $\preceq$ on infatoms, define the *information-theoretic downset* $\nabla^C_{\preceq}(\alpha)$ of a sentence α as follows: $\nabla^C_{\preceq}(\alpha) = \{i \mid \exists j \in C_{\preceq}(\alpha)$, such that $i \preceq j\}$. And now define a removal in terms of a faithful total preorder $\preceq$ on infatoms using the following identity:

(Def $\sim$ from $\nabla^C_{\preceq}$) $K \sim \alpha = Th(C(K) \setminus \nabla^C_{\preceq}(\alpha))$

An appeal to part (6) of proposition 2 and part (4) of proposition 3 shows that (Def $\sim$ from $\nabla^C_{\preceq}$) is just an infatom version of (Def $\sim$ from $\nabla^M_{\preceq}$). As a result, the removals defined in terms of the faithful total preorders on infatoms using (Def $\sim$ from $\nabla^C_{\preceq}$) are precisely the severe withdrawals.

Severe withdrawal can now be justified as follows. Like AGM contraction, as defined using (Def

$-$ from $C_{\preceq}$), (Def $\sim$ from $\nabla^{C}_{\preceq}$) ensures that withdrawing α from K results in the removal, from K, of the least entrenched content bits of α. But unlike AGM contraction, all the infatoms that are at most as entrenched as the least entrenched content bits of α are also removed. The reason for this last removal is that these infatoms are at most as entrenched as the least entrenched content bits of α, and thus ought to be easier to dislodge from K.

The systematic withdrawal of Meyer et al. [15] can be justified along similar lines. The only difference between severe and systematic withdrawal is that the latter uses a different class of faithful preorders in (Def $\sim$ from $\nabla^{M}_{\preceq}$). Levi [12] has suggested two forms of principled withdrawal, which was subsequently formalised by Hansson and Olsson [9]. Levi has recently criticised his own constructions and defined a form of withdrawal which he refers to as *mild contraction*. It turns out that, when placed in the same context, mild contraction coincides with severe withdrawal.

6 Infobase change

While AGM belief change serves as a very useful general framework for belief change, it is generally accepted that belief sets do not have a rich enough structure to serve as appropriate models for epistemic states [4]. One of the proposals to rectify this is to replace belief sets with arbitrary sets of sentences, known as bases. In one view [16], a base should be thought of as providing more structure to its associated belief set, which means that it can be used to determine the belief contraction operation associated with a base contraction operation (where the belief contraction operation $-$ associated with a base contraction operation $\ominus$ is defined as: $Cn(B) - \alpha = Cn(B \ominus \alpha)$). This is the view adhered to by Meyer et al. [14] in their definition of infobase change. They define an infobase as a finite set of sentences, with each sentence viewed as being obtained independently from the others. Taking AGM belief change as the general framework in which to operate, they present a method that uses the structure of an infobase B to determine which AGM contraction operation to associate with the infobase contraction operation to be constructed. Given an infobase B, they define the *B-number* v_B of a valuation v as the number of logically non-equivalent sentences of B (tautologies excluded) that are true in v. The B-numbers of the valuations are then used to obtain a $Cn(B)$-faithful total preorder $\preceq$ as follows: $v \preceq w$ iff $w_B \leq v_B$. In other words, the more sentences of B that are true in a valuation v, the lower down in the preorder $\preceq$, and the more plausible, it will be. It is easily verified that the relation $\preceq$ defined as such is indeed a $Cn(B)$-faithful total preorder. The total preorder $\preceq$ is then employed in the definition of an infobase contraction operation $\ominus$ by letting $-$ be the belief contraction operation defined in terms of $\preceq$ using (Def $-$ from $M_{\preceq}$), and requiring that $Cn(B \ominus \alpha) = Cn(B) - \alpha$. Although this is not enough to specify precisely what the resulting *infobase* should be when contracting B with a sentence α, it does specify the unique belief set to be generated by the resulting infobase. Meyer et al. do suggest a unique way for obtaining the resulting infobase, but the details need not concern us here.

By recasting their method into an information-theoretic framework, we obtain an appropriate justification for the use of these infobase contraction operations. Define the B-number i_B of an infatom i as the number of logically non-equivalent sentences in B of which i is a content bit (tautologies excluded), and then define the $Cn(B)$-faithful total preorder $\preceq$ in terms of the B-numbers of the infatoms as follows: $i \preceq j$ iff $i_B \leq j_B$. In other words, the number of times that an infatom occurs as a content bit of the sentences in B determines its entrenchment. Given the assumption of the independence of sentences in B, this is a very natural way in which to order infatoms. Consider the following example, which is an adaption of an example taken from [8]. I currently believe that Cleopatra had a son and a daughter. This is based on information obtained from two different history books; one asserting that Cleopatra had a son and the other that she had a daughter. I subsequently unearth information which leads me to retract my belief that Cleopatra had *both* a son and a daughter. Given the independence of the initial two claims it then seems reasonable to retain the belief that Cleopatra had a child (without claiming to know whether it is a boy or a girl). To formalise this example, let L be a propositional language generated by the two atoms p and q, and let p denote the assertion that Cleopatra had a son and q the assertion that she had a daughter. My current beliefs about Cleopatra can then be expressed in terms of the infobase $B = \{p, q\}$. Let Inf= $\{i, j, k, l\}$, where $i(p) = I$, $i(q) = I$, $j(p) = I$, $j(q) = E$, $k(p) = E$, $k(q) = I$, $l(p) = E$, and $l(q) = E$. That is, i is a content bit of both p and q, j is a content bit of p but not of q, k is a content bit of q but not of p, and k is a content bit of neither p nor q. It is easily verified that $i_B = 2$, $j_B = 1$, $k_B = 1$, and $l_B = 0$. That is, the most entrenched infatom is i, the one that is a content bit of both p and q. This is followed by j and k, the infatoms that are content bits of exactly one of p and q. And finally we have l, the infatom which is a content bit of neither p nor q, as the least entrenched infatom. Now, let $-$ be the AGM contraction defined in terms of $\preceq$ using (Def $-$ from $C_\preceq$), where $\preceq$ is the $Cn(B)$-faithful total preorder obtained from the B-numbers of the infatoms. It can easily be verified that $Cn(B) - p \wedge q = Cn(p \vee q)$. Using the method of Meyer et al. we would thus have that $Cn(B \ominus p \wedge q) = Cn(p \vee q)$, where $\ominus$ is an infobase contraction. In other words, retracting the information that Cleopatra had both a son and a daughter yields an infobase from which we can conclude that she had a child without claiming to know whether she had a son or a daughter (or both). And this corresponds exactly with the intuition set out above.

The crux of the matter is that the infatom i is viewed as the most entrenched infatom because it forms part of both of the independently obtained sentences in the infobase B. And as a result, the two remaining content bits of $Cn(B)$, the infatoms j and k, are removed when contracting by $p \wedge q$.

7 Conclusion

We have provided an information-theoretic semantics for AGM contraction and related constructions and operations in the field of belief change. Such a semantics provides new forms of justification for the use of these belief change operations. The underlying intuition is that of an ordering of entrenchment of the bits of information from which the current set of beliefs of an agent is built up; an idea which can be used to guide the development of new belief change operations as well.

References

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] Rudolf Carnap and Yehoshua Bar-Hillel. Semantic information. *British Journal for the Philosophy of Science*, 4:147–157, 1953.

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

[4] Peter Gärdenfors. *Knowledge in Flux : Modeling the Dynamics of Epistemic States.* The MIT Press, Cambridge, Massachusetts, 1988.

[5] Peter Gärdenfors and David Makinson. Revisions of knowledge systems using epistemic entrenchment. In M.Y. Vardi, editor, *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 83–95. Morgan Kaufmann, Los Altos, California, 1988.

[6] Peter Gärdenfors and David Makinson. Nonmonotonic inference based on expectations. *Artificial Intelligence*, 65:197–245, 1994.

[7] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[8] Sven Ove Hansson. Knowledge-level analysis of belief base operations. *Artificial Intelligence*, 82:215–235, 1996.

[9] Sven Ove Hansson and Erik J. Olsson. Levi contractions and AGM contractions: A Comparison. *Notre Dame Journal of Formal Logic*, 36(1):103–119, 1995.

[10] Sven Ove Hansson and Hans Rott. How not to change the theory of theory change: A reply to Tennant. *British Journal for the Philosophy of Science*, 46:381–383, 1995.

[11] H. Katsuno and A.O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[12] Isaac Levi. *The Fixation of Belief and Its Undoing.* Cambridge University Press, Cambridge, 1991.

[13] Isaac Levi. Contraction and informational value. Unpublished manuscript, Columbia University, February 1998.

[14] Thomas A. Meyer, Willem A. Labuschagne, and Johannes Heidema. Infobase Change: A First Approximation. *Journal of Logic, Language and Information (to appear)*, 1999.

[15] Thomas A. Meyer, Louise Leenen, Willem A. Labuschagne, and Johannes Heidema. Systematic withdrawal. Technical report, University of South Africa, 1999.

[16] Bernhard Nebel. A knowledge level analysis of belief revision. In Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, editors, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 301–311, San Mateo, CA, 1989. Morgan Kaufmann.

[17] Hans Rott and Maurice Pagnucco. Severe withdrawal (and Recovery). *Journal of Philosophical Logic (to appear)*, 1999.