Relevance Sensitive Belief Structures

Samir Chopra
CUNY Graduate Center
schopra@gc.cuny.edu

Rohit Parikh*
CUNY Graduate Center
ripbc@cunyvm.cuny.edu

September 26, 2000

Abstract

We propose a new relevance sensitive model for representing and revising belief structures, which relies on a notion of partial language splitting and tolerates some amount of inconsistency while retaining classical logic. The model preserves an agent's ability to answer queries in a coherent way using Belnap's four-valued logic. Axioms analogous to the AGM axioms hold for this new model. The distinction between implicit and explicit beliefs is represented and psychologically plausible, computationally tractable procedures for query answering and belief base revision are obtained.

*Research supported by NSF and the CUNY-FRAP program. The authors wish to thank Samuel Buss, Melvin Fitting, Konstantinos Georgatos, Bud Mishra, Graham Priest and Renata Wassermann for helpful comments. Part of this research was carried out while both authors were visiting New York University and we thank NYU for its hospitality. Portions of this study appeared in [6].

1 Introduction

When we consider the belief structures of ordinary people as contrasted with formal logical theories, some features of the former immediately stand out.

One feature is that people tend to divide their beliefs more or less roughly into *subject areas* and reason in actual situations with restricted subsets of their beliefs. Thus a person who is asked a question about French history will of course consult his beliefs about Napoleon, perhaps even about English or European history, but is unlikely to consult his beliefs about Indian music or the distance of the Andromeda galaxy from us.

A second feature is that the actual beliefs of people are typically inconsistent. It took a Supreme court decision to decide that segregation is incompatible with equality of human beings. Until such a decision took place in 1954, many Americans lived more or less comfortably with two incompatible beliefs and were able to go on with their lives, rather than, as a logical theory might, derive all possible formulas from two incompatible ones and have the reasoning process come to a dead halt.

A third feature is that human beings are not in practice deterred by the fact that derivability in the propositional calculus is co-**NP**-complete and that in the predicate calculus it is actually undecidable. It is common for people to answer a query with "I don't know", but not with "I don't know whether an answer follows from my beliefs". If computational complexity were a problem in real life, such answers would be common-place.

Clearly, it seems that, to model human reasoning processes (and the processes of human-like computing devices), we need to come to terms with these three facts. In the current paper we propose a new kind of structure which does just that.

In this paper we propose a new model – the B-structures model – for representing belief structures which relies on a notion of *partial language splitting* and will tolerate *some amount* of inconsistency while retaining classical logic *locally*. This model will aim to:

- Provide a representation for beliefs that is relevance sensitive and inconsistency tolerant.

- Provide psychologically plausible, computationally tractable procedures for belief revision.

- Provide a query answering scheme that preserves an agent's ability to answer queries in a coherent way from a possibly inconsistent epistemic state. Such a scheme will be particularly useful for large data bases.

- Respect the minimal change requirement for belief revision: when a belief structure is revised in the face of new information, we would like the new belief structure to be as similar to the old one as possible. This requirement goes back to Quine and has been repeatedly emphasized by Gärdenfors as the *preservation criterion* (cf. [Gärdenfors, 1988].

- Explicate the distinction between implicit and explicit belief: there are beliefs which are explicit - they may be actually asserted or at least agreed to. Other beliefs are implicit in the sense that an agent can be made to agree to them after a discussion. If the set of explicit beliefs is consistent, then we can define the set of implicit beliefs as just their logical consequences. However, this becomes implausible if – unknown to the agent – the explicit beliefs are inconsistent.

The B-structures model will extend the *splitting languages* framework of [29] by incorporating its intuitions and going beyond them. That framework showed, using a partly syntactic notion of relevance, how an agent's beliefs could be (uniquely) divided into sub-areas and how this division could be used in belief revision. It did not address the issue of explicit versus implicit belief, nor did it tackle the question of inconsistent beliefs. Here, we do both. We will represent an agent's beliefs, not by a *theory* as is usual, but by what we call a B-structure, which generalizes the notion of a theory. We begin by giving some notation and re-stating the celebrated AGM axioms [2] for belief revision.

Notation: In the following, L is a finite propositional language i.e., a set of propostional atoms. (However, most results continue to hold for a countably infinite first order language without equality.) The constants *true, false* are in L. We do not use the symbols $(\top, \bot)$ for truth and falsity but reserve them for an alternative interpretation under a four-valued logic (we will take $\bot$ to stand for the empty truth assignment i.e., *no information* and $\top$ to stand for the *over-defined* truth assignment). The letter $\mathcal{L}$ stands for the set of formulae generated by L. We will often refer to both these sets $L, \mathcal{L}$ as languages, but our notation will make it clear which one is being referred to. Greek letters $\alpha, \beta \ldots$ stand for arbitrary formulae, Roman lower case letters $p, q, r \ldots$ stand for propositional atoms. $\alpha \Leftrightarrow \beta$ means that $\alpha \leftrightarrow \beta$ is a tautology, i.e. true under all truth assignments. Similarly, $\alpha \Rightarrow \beta$ means that $\alpha \rightarrow \beta$ is a tautology. If X is a set of formulae then $Cn(X)$ is the logical closure of X. Thus X is a theory iff $X = Cn(X)$. Letters T, K possibly with subscripts denote theories. $T * \alpha$ is the revision of T by α, and finally, $T \dot{+} \alpha$ is $Cn(T \cup \{\alpha\})$, i.e. the result of a brute addition of α to T (followed by logical closure) without considering the need for consistency. Slightly departing from the AGM usage, we always assume that both T and α are *individually* consistent, only their union might not be.

AGM have proposed the following widely accepted axioms for the revision operator $*$. We omit their axioms[1] 7,8.

1. $T * \alpha$ is a theory.
2. $\alpha \in T * \alpha$
3. If $\alpha \Leftrightarrow \beta$, then $T * \alpha = T * \beta$.
4. $T * \alpha \subseteq T \dot{+} \alpha$
5. If α is consistent with T, i.e. it is not the case that $\neg\alpha \in T$, then $T * \alpha = T \dot{+} \alpha$.
6. $T * \alpha$ is consistent if α is.

Unfortunately the AGM axioms 1-6 are compatible with the *trivial update*[2] $*_t$:

Definition 1 *If α is consistent with T, then $T *_t \alpha = T \dot{+} \alpha$, otherwise $T *_t \alpha = Cn(\alpha)$.*

Thus in case α is inconsistent with T, under this update, all information in T is simply discarded, clearly something which is not in the spirit of the minimal change requirement. The work in [29], discussed below, is motivated in part by a desire to block the trivial update.

[1] See [29] for a discussion of these two axioms and why they are less plausible than the other six.
[2] Axiom 7 does block the trivial update. Since we do not accept the axiom, this is not an alternative for us.

2 The Language Splitting Model

Take an agent's epistemic state to be represented by a theory T in the finite propositional language L, and let $\{L_1, L_2\}$ be a partition of L. We then have the following definition of the *splitting* of a theory by languages:

DEFINITION 2 [29] *(i) (mutually disjoint) languages* L_1, L_2 split *the language* L *relative to a theory* T *iff there ar formulae* α, β *such that* $\alpha \in L_1, \beta \in L_2$, *and* $T = Cn(\alpha, \beta)$.
More generally, languages $L_1, \ldots, L_n$ *split* T *iff there exist* $\alpha_i \in L_i$ *such that* $T = Cn(\alpha_1, \ldots, \alpha_n)$. *We say that* $\{L_1, \ldots, L_n\}$ *is a* T*-splitting and that* T *is 'generated' by the* T_i *in the sub-languages* L_i.
(ii) Suppose that $L_1 \subset L$. *Then* T *is said to be* confined *to* L_1 *if* $T = Cn(T \cap L_1)$.

The first part of the definition implies that T contains no 'cross-talk' between L_i and L_j for distinct i, j. The second part says that T knows nothing about the part $L - L_1$ of L. The following lemmas are instrumental in proving the existence of a new update operator for theories split by language:

LEMMA 1 [29] *Given a theory* T *in the language* L, *there is a unique* finest T*-splitting of* L, *i.e., one which refines*[3] *every other* T*-splitting.*

Lemma 1 shows that there is a unique way to think of T as being composed of disjoint information about certain subject matters. We can think of a theory as being sliced up in many different ways; when put back together, each slicing gives us the original theory. However, there always exists a unique way of carving up the theory that is finer than any other. This lets us represent an agent's beliefs as the union of disjoint subjects so that we do not lump together subjects that could be treated separately. How, though, do we determine the subject matter of a belief in the theory T?

LEMMA 2 [29] *Given a formula* α *and a language* L, *there is a smallest language* L' *(a set of propositional atoms) in which* α *can be expressed, that is, there is a* $L' \subseteq L$ *and a formula* $\beta \in L'$ *with* $\alpha \Leftrightarrow \beta$, *and for all* L'' *(another set of propositional atoms) and* β'' *such that* $\beta'' \in L''$ *and* $\alpha \Leftrightarrow \beta''$, $L' \subseteq L''$.

Although α is equivalent to many different formulas in different languages, the lemma tells us that nonetheless, the question, "What is α actually *about*?" can be uniquely answered. The *subject matter* of a proposition is the *smallest language* of a formula used to express the proposition. The language L' will be referred to as L_α. If $\alpha \Leftrightarrow \beta$, then $L_\alpha = L_\beta$.

Example: If α is the formula $p \wedge (q \vee \neg q)$, then $\alpha \Leftrightarrow p$ and hence $L_\alpha = \{p\}$.

The following (not entirely syntactical) definition of relevance is used in the language splitting (henceforth **LS**) model:

[3] A partition P refines another partition P' if *every* element of P is a subset of *some* element of P'.

DEFINITION 3 *Formulas* α, β *are* relevant *to each other,* $\mathcal{R}(\alpha, \beta)$ *iff their* smallest *languages* L_α, L_β *share propositional atoms, i.e., iff* $L_\alpha \cap L_\beta \neq \emptyset$

Using the smallest language blocks trivial relevance between two formulas. For example, if we were not using the smallest language, the formulas $\alpha = p \wedge (q \vee \neg q), \beta = r \wedge (q \vee \neg q)$ would both be deemed relevant to q and to each other. But under our definition $L_\alpha = \{p\}$ and $L_\beta = \{r\}$ and so α, β are deemed irrelevant to each other. Note that it is possible that two formulas might be intuitively irrelevant to each other even if they share atoms and that that do not share atoms might, in some contexts, be thought of as being relevant to each other. Such judgements are entirely possible and it is not our intention to capture the full range of judgements of relevance by our definition[4].

We now clarify the notion of a language *compatible* with a given T-splitting:

DEFINITION 4 [29] *Given a theory* T*, a language* L*, and a formula* α*, let* L_α *be the smallest language in which* α *can be expressed and* L_α^T *be the smallest language containing* L_α *such that* $\{L_\alpha^T, L - L_\alpha^T\}$ *is a* T*-splitting.*

L_α^T is the smallest union of elements of the finest T- splitting of L in which α can be expressed.
Example: Suppose $T = Cn(p \vee q, r, s)$ and $\alpha = q \vee r$. Then T is split by the languages $\{\{p,q\},\{r\},\{s\}\}, L_\alpha = \{q,r\}$ and $L_\alpha^T = \{p,q,r\}$.

The Language Splitting Axioms

The motivation for the language splitting axioms is that when an agent has information about two subject matters which, as far as the agent believes, are unrelated, (are split) then when the agent receives information about one of the two, it should only update information in that subject and leave the rest alone. This central intuition is expressed by the following (which supplements AGM axioms) in the statements of the axioms below, all operators are intended to be general AGM-compatible update operators:
Axiom P: If $T = Cn(\alpha, \beta)$ where α, β are in L_1, L_2 respectively and γ is in L_1, then T updated by γ is $(Cn(\alpha) \oplus \gamma) \dot{+} \beta$, where $\oplus$ is an update operator for the language L_1.
Axiom **P** ensures that incoming information does not 'damage' existing information to which it is not relevant.
Three additional axioms for revision on theories split by language follow immediately:
Axiom P1: If T is split between L_1 and L_2, and α is an L_1 formula, then $T * \alpha$ is also split between L_1 and L_2.
Justification: The new information received makes no connection between the partitions in the theory, so after revision the old partition should be retained.
Axiom P2: If T is split between L_1 and L_2, α, β are in L_1 and L_2 respectively then $T * \alpha * \beta = T * \beta * \alpha$.
Justification: The order of the receipt of two unrelated pieces of information should make no difference to the revision of our beliefs.

[4]Note that the developments in this paper go back to work in 1996 which may be found in [29]. A detailed investigation of syntactic and semantic notions of relevance was presented in 1998 in [23]. A 1997 development of a probability inspired 3-place notion of relevance among sets of propositional atoms occurs in [8]. A comprehensive collection of papers dealing with relevance may be found in [1].

An axiom stronger than **P2** has been suggested by K. Georgatos:
Axiom P2g: If T is split between L_1 and L_2, α, β, are in L_1 and L_2 respectively, then $T * \alpha * \beta = T * \beta * \alpha = T * (\alpha \wedge \beta)$.
This axiom is shown in [29] to be compatible with **P**.
Axiom P3: If T is confined to L_1 and $\alpha \in L_1$ then $T * \alpha$ is just the set of consequences in L of $T \oplus \alpha$ where $\oplus$ is the update of T by α in the sublanguage L_1.
Justification: The agent possesses no information about $L - L_1$ and has not received any yet. If so, the agent should update as if it held beliefs only in L_1. $L - L_1$, about which we had no prior information and have not received any updates should not have any impact.

OBSERVATION 1 [29] *The trivial update*[5] *(as in Definition 1), cannot satisfy Axiom* **P2** *or Axiom* **P** *though it does satisfy* **P1** *and* **P3**. *Therefore, any update operator that satisfies Axiom* **P** *cannot be the trivial update.*

Demonstration: Let $T = Cn(p, q)$ and consider new inputs $\alpha = p$ and $\beta = \neg q$. Then (since $T*\alpha = T$) the trivial update yields $T*\alpha*\beta = T*\beta = Cn(\neg q)$ and $T*\beta*\alpha = Cn(\beta)*\alpha = Cn(\neg q, p)$. This violates **P2**. Also $T * \beta = Cn(\neg q)$ which violates Axiom **P**. Therefore **P** or **P2** alone rules out the trivial update.
The following theorem provides a new update operator for theories[6] on which splittings can be carried out:

THEOREM 2.1 [29] *There exists an update operator that satisfies Axiom* **P** *and axioms* $(K^*1) - (K^*6)$*; the update operator blocks the trivial update.*

One *possible* revision procedure is defined as follows:

DEFINITION 5 [29] *Given* T, α, *if* $\neg\alpha \notin T, T *_s \alpha = T \dot{+} \alpha$.
Otherwise recall that T *can be expressed as* $Cn(\beta, \gamma)$ *where* $\beta \in L^T_\alpha, \gamma \in L - L^T_\alpha$ *respectively. Now let* $T *_s \alpha = Cn(\alpha, \gamma)$.

Note that we used the trivial update locally, i.e., in $L - L^T_\alpha$. But this is done for convenience, and another local update which retained more information could have been used.
Example: Let $T = Cn(p, q \vee r, s)$; the partition $\{\{p\}, \{q, r\}, \{s\}\}$, is the finest T-splitting. Let $\alpha = \neg p \wedge \neg q$. $L_\alpha = \{p, q\}$. Since T connects q and r, we get $L^T_\alpha = \{p, q, r\}$. Now let $\beta = p \wedge (q \vee r)$ and $\gamma = s$. β is the part of T in L^T_α and also the part of T incompatible with the new information α. However, γ has no quarrel with α and is untouched by it. Therefore $T *_s \alpha$ is $Cn((\neg p \wedge \neg q), s)$. The update procedure of the theorem notices that α is not in any conflict with γ and retains the latter. It should be obvious to the reader that other revision procedures exist.

[5] If α is consistent with T, then $T * \alpha = T \dot{+} \alpha$, otherwise $T * \alpha = Cn(\alpha)$.

[6] The notion of splitting languages and the lemmas can be extended to first order logic without identity (the introduction of identity introduces problems pointed out by Sam Buss). Arguments similar to those in [29] can be provided to show the existence of an update operator satisfying AGM axioms and Axiom **P** for theories in first-order languages. Furthermore, the **LS** model is not bound to a particular assignment of propositional atoms to the language. As an example, the language $L = \{p, q\}$ is also generated by r, q where r is $p \leftrightarrow q$. This follows immediately from the fact that p can be expressed in terms of q, r as $p \Leftrightarrow (q \leftrightarrow r)$.

The language splitting model provides us with substantial computational gains since the size of the set to be updated is always *a fraction* of the size of the original set and ensures the following desirable result:

> [W]hen we get a piece of information which lies in one of the sub-languages (or straddles only two or three of them) then we can leave most of the theory unchanged and revise only the affected part.[7]

We now have a tractable, localized, relevance-sensitive belief revision to work with. The minimal change requirement is satisfied by using relevance relations amongst beliefs to consider as small a part of the old theory as possible as a candidate for change. This ensures that while the complexity of the revision procedure stays at the same level, the size of the problem instance is drastically reduced.

3 Relevance Relations for Belief Revision

The **LS** model provides a relevance sensitive model for belief revision. It is this sensitivity that we will exploit in the models that we develop. The problem of defining a notion of relevance between propositions has been notoriously tough; defining a notion of relevance that is adequately formalizable will help in demonstrating how changes in our beliefs do not seem to necessitate widespread changes in the rest of our beliefs.

The language splitting model splits an agent's beliefs into a number of 'component' theories and the only changes made during belief revision are to component theories that are in the same language or have language overlap with the new information[8].

Language overlap as a relevance relation as used in the language splitting model is intuitively more plausible than other axiomatizations of relevance relations in the literature. For example, Epstein [10] places the following conditions on an arbitrary relevance relation $\mathcal{R}$ (for propositions α, β):

- **R1:** $\mathcal{R}(\alpha,\beta)$ iff $\mathcal{R}\ (\neg\alpha,\beta)$
- **R2:** $\mathcal{R}(\alpha,\beta \wedge \gamma)$ iff $\mathcal{R}\ (\alpha,\beta \rightarrow \gamma)$
- **R3:** $\mathcal{R}(\alpha,\beta)$ iff $\mathcal{R}(\beta,\alpha)$
- **R4:** $\mathcal{R}(\alpha,\alpha)$
- **R5:** $\mathcal{R}\ (\alpha,\beta \rightarrow \gamma)$ iff $\mathcal{R}\ (\alpha,\beta)$ or $\mathcal{R}(\alpha,\gamma)$

Epstein intends the above as conditions on *any* plausible notions of relevance; some are more intuitive than others. For instance it is not clear what intuition about relevance **R2** captures. If

[7] [29], page 5.

[8] This is analogous to the corresponds to the *sleeping dog strategy* suggested by McDermott [24] as a solution for the frame problem:

> [AI systems] to reason about result (s,e) (i.e., about the result of event e in a situation) compute all the effects of e in situation s, make those changes, and leave the rest of s (the sleeping dogs above) alone.

we let $\beta = p, \gamma = \neg p$, then $\beta \to \gamma$ is equivalent to $\neg p$ and of course relevant to p. However, $\beta \wedge \gamma$ is $p \wedge \neg p$, a downright contradiction and not relevant (in our opinion) to *anything*. This fact casts some doubt on the intuition behind **R2**. Indeed it turns out that **R2,R5** are incompatible with the following natural requirement:

R6: If $\beta_1 \Leftrightarrow \beta_2$ then $\mathcal{R}(\alpha, \beta_1)$ iff $\mathcal{R}(\alpha, \beta_2)$.

We make a distinction here between $L(\alpha)$, the set of propositional variables which actually occur in a formula α and the actually germane L_α as in Definition 3. We thus distinguish between two possible relevance relations based on language overlap:

DEFINITION 6 *α, β are related by* syntactic *language overlap ($\mathcal{R}_s(\alpha, \beta)$) if $L(\alpha) \cap L(\beta) \neq \emptyset$. α, β are related by* logical *language overlap ($\mathcal{R}_l(\alpha, \beta)$) if $L_\alpha \cap L_\beta \neq \emptyset$.*

It is obvious that $\mathcal{R}_l(\alpha, \beta))$ implies $\mathcal{R}_s(\alpha, \beta))$. Rodrigues [32] has shown that the relation $(\mathcal{R}_s(\alpha, \beta))$ is the smallest relation satisfying Epstein's conditions. However:

OBSERVATION 2 *The relevance relation $\mathcal{R}_l(\alpha, \beta)$ using the smallest languages for α, β satisfies conditions* **R1, R3-4, R6** *above. It does not satisfy conditions* **R2, R5**. *However, it does satisfy* half *of* **R5***:*
R5a: *If $\mathcal{R}_l(\alpha, \beta \to \gamma)$ then $\mathcal{R}_l(\alpha, \beta)$ or $\mathcal{R}_l(\alpha, \gamma)$*

The definition of relevance provided in the **LS** model let us define the subject matter of a proposition as the language of the formula used to express the proposition. This lets us speak of subject matters being identical or overlapping. That is, for propositions α, β, if $L_\alpha = L_\beta$ then α, β are about the same subject; if $L_\alpha \cap L_\beta \neq \emptyset$ then the two propositions have some subject matter in common and if $L_\alpha \subseteq L_\beta$ then the subject matter of α is contained in the subject matter of β.

4 The B-structures Model

The **LS** model requires that an agent's beliefs be partitioned into theories which have totally disjoint languages. In practice, however, beliefs in different areas *do* have some overlap in subject matter and so the partition of the main language is not actually strict. An agent's component theories do contain beliefs that are more relevant to one another than to beliefs in other component theories, but they are not *totally* irrelevant to beliefs in other subtheories.

Consider the following example: Bill Clinton's problems over Monica Lewinsky may perhaps be connected with the bombing of Iraq, which in turn may affect the price of oil and which may then affect my airfare to India. Normally, we do not reason with all these topics together at one time. More often than not, they will be kept separate. However, the connections may be noticed on occasion, if there is an article in the newspaper connecting a price hike in airfares with the shortage of oil.

Thus what we propose is a modification of the **LS** model in which distinct sublanguages may overlap; it may be the case that $L_i \cap L_j \neq \emptyset$ for distinct i, j. We also want it to be the case that each of the subtheories is individually consistent. Yet, we are now able to allow for the possibility of the global inconsistency of the belief state, something that could not be represented in the **LS** model.

Suppose an agent believes theories T_i in languages L_i and the L_i are mutually disjoint. Then if the T_i are individually consistent, they are also jointly consistent. Thus the **LS** model cannot explain how an agent can be locally logically omniscient – i.e., derive logical consequences *within* each L_i but still fail to be globally consistent. But the B-structure model we shall now propose will permit agents to be globally inconsistent (as unfortunately most of us are) while still operating locally in consistent frameworks.

Note that the sort of model we are considering will also be appropriate for group reasoning. Each agent will have its own language and be individually consistent in it. But the languages of different agents may overlap and a particular question α may be resolved by consulting those agents whose languages overlap with the language of α. The revision sch discuss in the following pages will show how this is handled at a technical level. Even a single agent may be thought of as consisting of a collective entity, see [25].

Our approach towards inconsistency will be distinct from paraconsistent approaches in that we will allow full use of classical logic, locally, while using a multi-valued logic for answering queries. Since the agent's beliefs can be inconsistent, we will need to keep the B-structure logically open and define a companion notion of inference in order to define a coherent set of implicit beliefs. This inference operation will block the derivation of outright inconsistencies from the agent's (possibly inconsistent) belief representation.
Three basic definitions are all we need to establish the B-structure model.

DEFINITION 7 *A* belief structure $\mathcal{B}$ *on* L *is a set* $\{(L_1, T_1), ..., (L_n, T_n)\}$ *such that* $L = \bigcup L_i : i \leq n$, *and each* T_i *is a consistent, finitely axiomatizable theory in* L_i*. The* T_i *are* $Cn(\Gamma_i)$ *where the* Γ_i *are the explicit beliefs of the agent in language* L_i*. If* $n = 1$*, then* $\mathcal{B}$ *will be just a theory.*

A B-structure is a collection of sets of beliefs[9], each one of which is individually closed and consistent. Their union *may* be inconsistent, however, and querying, a sort of Socratic prodding, can expose this inconsistency under the right conditions. The requirement of maintaining a B-structure as a set of consistent theories captures the intuitive feeling that most of us have of being 'reasonably' rational agents. In so thinking, we take ourselves to be aware of the beliefs implicit in our explicit beliefs. Yet, these tend to be just a small subset of our total set. Being asked to consider whether we are aware of our implicit beliefs normally leads us to evaluate one theory at a time. Considering larger and larger portions of our beliefs can force us to drop our earlier claims of being aware of all the consequences of our beliefs.

Since we want to retain *some* amount of disjointness among the L_i, the notion of a k-partition captures the notion of limited overlap:

DEFINITION 8 *Given a finite propositional language* L *and a set of languages* $\{L_1 \ldots L_n\}$ *such that* $L = L_1 \cup \ldots \cup L_n$, $\{L_1 \ldots L_n\}$ *is a* k-partition *of* L *if any propositional symbol* q *occurs in at most* k *of the languages* L_i*.*

The overlap of symbols may be thought of as "cross-talk" amongst the T_i; beliefs in some of the T_i may be relevant to beliefs in others. A k-partition with a smaller k will be *more* disjoint

[9] A distinction should be made here with the notion of *mixed representations* as used by Nayak [26]. A mixed representation is the closure of more than one set and is not identical to what we have suggested here.

than one with a larger k and indicates a finer partitioning of her beliefs by the agent. In particular, a normal partition is just a 1-partition. A k-partition is *a fortiori* a $k+1$-partition and hence, for instance, a maximally fine 2-partition will usually be finer than the finest 1-partition whose existence is guaranteed by Lemma 1. Tolerating some overlap makes it easier to organize one's beliefs in smaller chunks.

Example: If T is axiomatized by the formula $(p \vee q) \wedge (\neg q \vee r)$ then $L(T) = \{p, q, r\}$ has only a trivial 1-partition. However, there is a 2-partition into the sets $\{p, q\}$ and $\{q, r\}$ with the two theories being generated by $p \vee q$ and $\neg q \vee r$ respectively.

Given the definition of relevance (see Definition 3) carried over from the language splitting model, the motivation for the definitions of the B-structures model is clear. If a B-structure is in languages L_1 and L_2, and $\alpha \in T_1$ then there might also be a formula $\gamma \in T_2$ such that $L_\alpha \cap L_\gamma \neq \emptyset$: α is relevant to γ. Using this approach, we want to avoid the *trivial relevancy* which arises when all beliefs are relevant to one another. It is precisely this situation that leads to the trivial update in AGM update operations: when we consider no language partition to exist. Furthermore, if B consists of T_1 and T_2, and $\alpha \in T_1$, then we suppose that for any $\beta \in T_1$, and any $\gamma \in T_2$, α is more relevant to β than to γ and β is more relevant to α than to γ. We place beliefs in subsets according to how relevant we take them to be to one another: this is a subjective placement, dependent on the agent. We might inspect an agent's B-structure and find that objectively, a different splitting could have been carried about (because of the languages involved).

The following definition captures how coherent an agent's beliefs can be while being inconsistent:

Definition 9 *A belief structure $\mathcal{B}$ as in Definition 7 is m-*consistent *iff any m of the T_i are jointly consistent.*

We do not require that the whole collection be consistent, but the *amount* of inconsistency is limited by this requirement. If $\mathcal{B}$ is 1-consistent, that merely says that all the T_i are individually consistent. At the other extreme, if $\mathcal{B} = \{(L_1, T_1), ..., (L_n, T_n)\}$ is n-consistent, then it will be consistent, period. The level of m-consistency is a measure of the coherence of the agent's beliefs[10].

In between 1-consistency and n-consistency lies an entire gradient of levels of inconsistency; between the thoughtless and the omniscient lies all of humanity. Most human agents might acknowledge their inconsistency in some matters and, yet, take themselves to be consistent in most of their reasoning. As long as we do not put too many of our theories together at one time, we retain consistency. It is this aspect of reasoning that m-consistency expresses. If m is suitably large, though less than n, the agent can operate in most contexts.

Examples of B-structures

The Bald Person: A natural example of such a collection arises in the context of the *bald person paradox*. Suppose we say that a person with 10,000 hairs is not bald but that a person with 0 hairs *is* bald. We also say that a non-bald person cannot become bald through the removal of just one hair. This gives us the axioms b_0, $b_n \rightarrow b_{n+1}$, for $n = 0$ to $n = 9{,}999$, and finally, $\neg b_{10{,}000}$. These 10,002 axioms are inconsistent, but any 10,001 of them are consistent. This gives us a 10,001-consistent,

[10] A similarity may (with similar motivation) be noted here with Kyburg's [22] *Principle of n-wise Consistency*: If K is a body of reasonably accepted statements, then for no set of statements $s_1, s_2, s_3 \ldots s_{n-1}$ is it the case that each of $s_1, s_2, s_3 \ldots s_{n-1}$ and $\neg(s_1, s_2, s_3 \ldots s_{n-1})$ is a member of K.

inconsistent collection of axioms, all of which are accepted in practice. They could be collected together in to 10,002 component theories whose axioms are b_0, $b_i \rightarrow b_{i+1}$ for $i < 10,000$, and finally $\neg b_{10,000}$. This example suggests that we can coherently speak of B-structures being formed with respect to particular situations and contexts; in this situation, a B-structure that contains our beliefs about the baldness predicate and avoids the bald man paradox through insulation.

The Wily Airline: For a more practical example, consider an airline which has 100 seats on a flight, and accepts 110 reservations. For $i \leq 110$ let c_i express the condition that customer number i has a seat. The airline asserts to customer i that c_i is true. A gigantic formula α expresses that at most 100 of the c_i *can* be true. This is an inconsistent set of assertions, organized into a B-structure with 111 component theories. The inconsistency will not matter if 100 or fewer of the customers actually show up at the gate and demand a seat.

A similar situation arises with a bank which guarantees each of its depositors that it will pay him/her the entire balance on demand, but in fact has cash on hand to pay only a minority, the remaining funds being in stocks or home mortgages. Any particular customer who demands his money will get it, but by the very nature of things they cannot *all* get their funds held by the bank. If there is no run on the bank, things will be fine. Again we have an inconsistent but eminently useful B-structure which makes bank customers feel secure while releasing money for people to buy houses with.

A Four Valued Semantics

To represent how an agent *uses* its B-structure and revises it, we make use of a four valued semantics provided by multivalued logics such as those devised by Belnap [3] and Fitting [11, 13] where the four truth values are $\perp$ (which stands for *no information*), *true, false*, $\top$ (which stands for *over-defined* or inconsistent). In Belnap's logic $\mathcal{FOUR}$, truth values can be thought of as *sets* of ordinary truth values. $\perp$ for $\emptyset$ (take it to indicate a lack of information), $true$ for $\{true\}$, $false$ for $\{false\}$ and $\top$ for $\{false, true\}$. In this four valued lattice there are two *partial orderings.*

The *knowledge* ordering ($<_k$) : In this ordering $\perp <_k$ *true* and *false* $<_k \top$. Generally, if α is below or the same level in the ordering as β, we can write $\alpha \leq_k \beta$. *true, false* are *incomparable* under the knowledge ordering. In the knowledge ordering, an increase means that at least one or more of the propositional atoms gains an assignment of *true* or *false* that it did not have, without losing any previous truth assignments that it previously had. The knowledge ordering represents degrees of information; so obviously, having no information, $\perp$, belongs at the bottom of the ordering and, having too much information, $\top$, is at the top.

The *truth* ordering ($<_t$) : In this ordering *false* $<_t \perp, \top <_t$ *true. true* is truer than $\top$ which is truer than *false.* Generally, if α is to the left of or the same as β, we can write $\alpha \leq_t \beta$ (β is at least as true as α is, and α is at least as false as β is). $\perp, \top$ are incomparable under the truth ordering. In the truth ordering, an increase means that one or more of the propositional atoms loses an assignment of *false* and/or gains one of *true.* The truth ordering corresponds to degrees of truth: so, $false \leq_t \top$ since $\top$, which corresponds to both true and false, is at least truer than $false$. Similarly, $\top \leq_t true$ since *true* is less false than $\top$, which is true *and* false.

The four truth values with their orderings are displayed in the double Hasse diagram below.

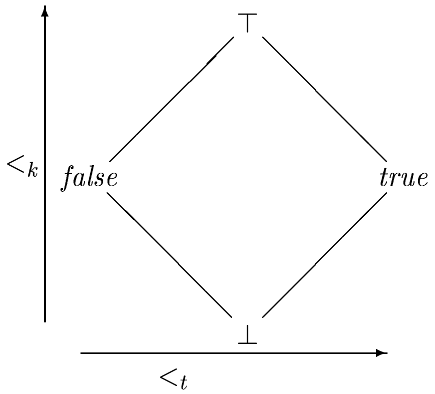

Fig. I, The Logic $\mathcal{FOUR}$

The motivation for picking a four valued logic to model queries is the same as that for picking a four-valued logic to represent the semantics of logic programs:

> Doing so provides a natural framework for the semantic understanding of logic programs that are *distributed over several sites, with possibly conflicting information coming from different places.*[11]

Now replace the phrase 'logic program' with belief structures, and the application of such a truth value space to the answers given by a possibly undecided, partly ignorant, partially omniscient, resource bounded agent (i.e., a realistic agent) becomes apparent.

5 Answering Queries

Our model shows that an agent whose beliefs are represented by a B-structure is usually able to answer questions coherently. An agent with an inconsistency present in its beliefs will still not respond in the affirmative to every query. We now provide a formalization of the intuitions at hand. The query "α?" is answered by giving the value $v_{\mathcal{B}}(\alpha)$ – abbreviated as $v(\alpha)$ – which is defined as follows:

DEFINITION 10 *Given a formula α and B-structure $\mathcal{B} = \{(L_1, T_1) \ldots (L_n, T_n)\}$ (each $T_i = Cn(\Gamma_i)$), let L_α be the smallest language in which α can be expressed. Now, let $\Gamma_\alpha = \bigcup\{\Gamma_i : L_i \cap L_\alpha \neq \emptyset\}$.*
If Γ_α is consistent, then
if $\Gamma_\alpha \vdash \alpha$, then $v(\alpha)$ = true
if $\Gamma_\alpha \vdash \neg\alpha$, then $v(\alpha)$ = false, and
$v(\alpha) = \bot$ otherwise.
If Γ_α is inconsistent, then $v(\alpha) = \top$.

Thus Γ_α is the union of those sets of explicit beliefs whose languages overlap with the language L_α of the query. Intuitively we see *which* of the theories T_i *could be* relevant to α and put them together (via the Γ_i) to get Γ_α. Γ_α is used to answer questions about α and the rest of the theories T_j are *not* brought into play. We can think of the Γ_α as the *evidence* for a particular proposition.

[11] Fitting, [12].

The more propositional symbols a formula has, the larger the number of parts of the B-structure which will be activated since L_α will overlap many languages in the B-structure.

In a court trial a judge will sometimes admonish the attorneys not to go on 'fishing expeditions'. Apart from a concern for privacy, the admonition also expresses a desire not to bring in too many 'component theories' of the relevant B-structure and thereby waste the time of the court in useless computation.

Limited Consistency

The following result states that under certain conditions of coherence of a B-structure, overlap of languages and the length of the query, the agent is guaranteed to give a consistent answer:

THEOREM 5.1 *Suppose that the collection $\{T_i : i \leq n\}$ is m-consistent. Then if L_α has at most l propositional symbols, the L_i are a k-partition, and $m \geq k \times l$, then Γ_α (as in Definition 10) will be a union of at most m of the theories T_i. Thus Γ_α will be consistent and exactly one of the first three values will be given.*

Proof: Immediate from definitions.

Therefore, an agent whose B-structure is *fairly consistent*, and who is responding to a *short* query, will give a consistent response. If I have beliefs about a few different subject matters, and I have these organized into groups of sub-theories each of which is individually consistent, and I have a high value of m for m-consistency, then, given that I respond to a query that only activates a small number of theories, the chances are high that I will come up with a consistent response.

An analogy to the query answering of an agent that organizes its beliefs in a B-structure is to think about the agent and its questioner as engaged in a *prover-adversary* game[12]. In that game the agent asserts that there is a truth assignment t that satisfies some long, inconsistent formula α. Since there is no such truth assignment, the prover asks the adversary the truth values under t of various formulas. The adversary loses if it assigns *true* to say $\alpha \wedge \beta$ but *false* to $\alpha \vee \beta$ or makes any other such error which is in conflict with the truth tables of the various logical connectives. Since such a truth assignment cannot exist, the agent will ultimately be caught in an inconsistency. However, if the prover is allowed only a *bounded* number of questions, the agent can get away with its concealed inconsistency. The following example is due to Pudlak and Buss [30]:
The Prover-Adversary Game: The formula $(\neg)^n(\alpha\wedge\neg\alpha)$ (with n even) requires $\log(n)$ questions to prove the agent's inconsistency. Suppose $n = 2^k$ and the prover asks the adversary the truth value of $(\neg)^{2^{k-1}}(\alpha \wedge \neg\alpha)$. If the agent gives the truth value of *true* (which is the wrong value) then n has been reduced by half and the prover can now work on the formula $(\neg)^{2^{k-1}}(\alpha \wedge \neg\alpha)$. If the adversary had instead given the value *false*, then there is an inconsistency between the truth value *false* for $(\neg)^{2^{k-1}}(\alpha \wedge \neg\alpha)$ and the truth value *true* for $(\neg)^{2^k}(\alpha \wedge \neg\alpha)$. The distance between the two formulas in terms of negation is now 2^{k-1}. Each time the prover spends one question, the distance between a correct value and an incorrect one is reduced by a factor of two. After at most $k = (\log n)$ steps, the distance will be one, and the agent will be caught.

The query answering procedure establishes the following set of *implicit beliefs* for an agent with a B-structure:

[12] Pudlak and Buss, [30].

Definition 11 *An agent with belief structure* $\mathcal{B}$ implicitly believes *the proposition* α *iff* $v_{\mathcal{B}}(\alpha) = true$. *Therefore, the set of implicit beliefs of the agent is the set of formulas which receive the valuation 'true' under the query answering scheme (as in Definition 10)* $= \{\gamma | v_{\mathcal{B}}(\gamma) = true\}$

If α is implicitly believed, then it follows from the agent's explicit beliefs, but the converse does not hold in general. If the explicit beliefs are jointly inconsistent, then their consequences will be the absurd belief set, but *the agent may still have implicit beliefs which are a reasonable set.* The implicit beliefs will always include the explicit ones and will be *locally* closed under logical rules.

The Fate of Adjunction

Is the rule of adjunction ($\alpha, \beta \vdash \alpha \wedge \beta$), preserved? If an agent implicitly believes α and β, does it also implicitly believe $\alpha \wedge \beta$? The answer provided by the B-structure model is sensitive to the *languages* in which the formulas α and β are expressed and to the *length* of the formulas. It may happen that a conjunction of two formulas α and β brings out an inconsistency which neither brought out by itself. Thus we may get the answer *true* to both α and β individually, but get $\top$ to the conjunction.

Example: Let $L_1 = \{p, r\}$, $L_2 = \{q, r\}$, $T_1 = Cn(p, r)$ and $T_2 = Cn(q, \neg r)$. Then the answers to the queries 'α?' ($\alpha = p$) and 'β?' ($\beta = q$) will both be *true*, but to '$\alpha \wedge \beta$?' it will be $\top$. The query "α?" only activates T_1, the query "β?" only activates T_2, but the query "$\alpha \wedge \beta$ activates both T_1 and T_2.

We now have a *principled reason* to reject the unrestricted application of the rule of adjunction. We might not accept the conjunction of (believed) statements that are large in number since there is a high probability that forming the conjunction of those statements exposes associated inconsistencies. This is, in essence, the bald man paradox.

Familiar puzzles such as the Lottery Paradox [21] and the Preface Paradox indicate the failure of the principle of adjunction in epistemic contexts. This epistemic principle can be expressed as the following (equivalent) statements:

> *The Conjunction Principle:* If S is a body of reasonably accepted statements, and s_1 belongs to S and s_2 belongs to S, then the conjunction of s_1 and s_2 belongs to S.
> *The Conjunctive Closure Principle:* If S is a body of reasonably accepted statements, then the conjunction of any finite number of members of S also belongs to S.

If the set S above is the set of implicit beliefs that are derivable from a B-structure, then this principle fails to hold in general.

The Lottery Paradox

Given a million lottery tickets, we believe, l_i with high probablitiy, where for each $i < 1,000,000, l_i$ = *Lottery ticket i will not win.* But if we believe that the lottery is fair then we also believe $\neg(l_1 \wedge l_1 \ldots \wedge l_{1,000,000})$. These 1,000,001 statements are jointly inconsistent, but any 1,000,000 of them are consistent. This gives us a 1,000,000-consistent, inconsistent collection of axioms, all of which are accepted in practice. The higher the value of m-consistency, the lesser the impression of paradox.

The Preface Paradox

In the Preface Paradox similarly, we have a set of beliefs, $p_1, p_2, \ldots, p_n$, which we believe individually and yet, it is also the case that we do not believe in their conjunction. Given a book with 500 pages, $p_1 \ldots p_n$ are all asserted where p_i = *page i does not contain an error.* But since the preface modestly asserts that there no doubt *are* errors in the book, $\neg(p_1 \wedge p_2 \ldots p_{500})$ is also asserted. These 501 statements are jointly inconsistent, but any 500 of them are consistent. This gives us a 500-consistent, inconsistent collection of axioms, which we find reasonable to work with in practice.

Sorites Paradoxes

Consider the following set of axioms:
I. 0 is small.
II. For all n, if n is small, then $n+1$ is small.
III. $10^{10^{10}}$ is not small.

Parikh [28] has shown that while the system **Peano Arithmetic** with axioms I-III above is *inconsistent*, all theorems proved in it whose proofs are *short* (and are purely about numbers) are *true*. The B-structures model provides a similar plausible treatment of Sorites type paradoxes as seen in the *bald person paradox* presented above.

In sum, the B-structures model makes the pragmatic claim that reasoning with collections of premises taken from a belief representation such as a B-structure is not likely to implicate us in embarrassing lines of reasoning.

6 B-Structure Refinement

An agent that has explicit beliefs Γ_i in language L_i may organize these beliefs in smaller or larger subtheories. If the Γ_i are inconsistent, then we cannot organize them into a single theory[13]. Organizing a B-structure into large subtheories makes computational problems harder since a more exhaustive search is required. It does have the advantage that more implicit beliefs can be derived. Organizing in smaller theories however, can mean a finer quarantining of inconsistencies and a lesser chance of being caught in inconsistency when reasoning with multiple theories. To formalize different strategies for compartmentalization of beliefs, we introduce the notion of a *refinement* of B-structures. In the definition below, we remind the reader that each theory $T_i = Cn(\Gamma_i)$ in a B-structure:

DEFINITION 12 *A B-structure* $\mathcal{B} = \{(L_1, T_1) \ldots (L_n, T_n)\}$ refines *another,* $\mathcal{B}' = \{(L'_1, T'_1) \ldots (L'_k, T'_k)\}$ *if*
(i) every language L_i *is a subset of some* L'_j,
(ii) every $L'_j = \bigcup\{L_i | L_i \subseteq L'_j\}$, *and*
(iii) every $T'_j = Cn(\bigcup \Gamma_i | L_i \subseteq L'_j)$.

A B-structure that refines another corresponds to a finer subdivision of the agent's beliefs.

[13]This is not the only reason, since as we have emphasized, splitting of beliefs is largely a resource conscious strategy.

Example: Suppose that the B-structure $\mathcal{B}$ has the *three* languages $\{p,q\}$, $\{q,r\}$, $\{r,s\}$ and theories $T_i : i \leq 3$ generated by the formulae $p \rightarrow q$, $\neg q \wedge r$ and $r \wedge s$ respectively, the B-structure $\mathcal{B}'$ has the *two* languages $\{p,q,r\},\{r,s\}$ and theories $T'_j : j \leq 2$ generated by the formulae $\neg p \wedge \neg q \wedge r$ and $r \wedge s$ respectively. Clearly $\mathcal{B}$ refines $\mathcal{B}'$. Now to the query 'p?' , $\mathcal{B}$ will give the answer $\perp$ and $\mathcal{B}'$ will give the answer 'no' or *false*. The B-structure $\mathcal{B}'$ has larger and fewer chunks than $\mathcal{B}$.

The precise relationship between the notion of refinement and the ranking of the queries in the four-valued logic of responses (more specifically, the ranking of queries on the knowledge ordering $<_k$ in the logic $\mathcal{FOUR}$) is provided by the following theorem:

THEOREM 6.1 *Let $\mathcal{B}$ refine $\mathcal{B}'$ and α be a formula. Then $v_{\mathcal{B}}(\alpha) <_k v'_{\mathcal{B}'}(\alpha)$.*

Proof: Given a formula α and B-structure $\mathcal{B}$ the query "α?" is answered by giving the value $v_{\mathcal{B}}(\alpha) = v(\alpha)$ which is defined using $\Gamma_\alpha = \bigcup \Gamma_i : L_i \cap L(\alpha) \neq \emptyset$, where $L(\alpha)$ is the language of α. Similarly for $\mathcal{B}'$. Now we note that if $L_i \subseteq L'_j$ and $L(\alpha)$ intersects L_i then it also intersects L'_j. Hence $\Gamma_\alpha \subseteq Cn(\Gamma'_\alpha)$. This immediately yields $v(\alpha) <_k v'(\alpha)$. $\Box$

In other words, every answer that $\mathcal{B}'$ provides to a particular query will rank higher on the knowledge ordering $<_k$ than the one provided by $\mathcal{B}$. However, $\mathcal{B}'$ may give the inconsistent answer $\top$ to a query α whereas $\mathcal{B}$ may have given a *true* or *false* answer to it. Then the answer from $\mathcal{B}'$ will be more 'knowledgable' but useless! The partitioning of information into separate languages may on occasion miss some answers which might have been obtained without such partitioning, but it ensures a greater likelihood of consistency in particular queries.

7 B-Structure Revision

There are two strategies available to an agent for handling belief revision on belief structures : generally, the sub-theories in which revision takes place will comprise all subtheories in which we find beliefs relevant to new information (the actual strategy followed will differ depending on context). In each case, the revision procedure will be sensitive to the smallest language that the new epistemic input is expressed in, letting us select the parts of B-structure that change on receipt of new information.

Strategies for Revision

Option A: Revise all subtheories relevant to new information without merger; in this method, every theory deemed relevant to a new formula will be revised individually. The agent accepts the consequences *within* each sub-language while still keeping them separate. If I learn that both Beijing and London had cold winters, I am not likely to merge all my other beliefs about the two cities.

Option B: Revision which involves the merger of theories: in this method, theories that are connected by an incoming piece of information are merged and revision is then carried out on the new theory so formed. This is likely to occur to a B-structure when an agent *repeatedly* receives some information which overlaps two sub-languages. The agent may decide that the division is artificial and should be abandoned.

7.1 Option A - The Non-merging Option

Assume that each of the languages L_i in a B-structure has its own AGM style revision operator $*_i$. We begin by defining the impact a new piece of information has on each sub-theory of the agent. We assume for simplicity that each L_i is a finite propositional language.

DEFINITION 13 *Let a belief structure $\mathcal{B} = \{(L_1, T_1), ..., (L_n, T_n)\}$ be given. Let α be a new input. Then, for each i, let the i-shadow of α be the set $\{\beta | \beta \in L_i \wedge \alpha \vdash \beta\}$. This i-shadow[14] will be a theory T_i' in L_i. T_i' is what α has to say in the language L_i. Let $\alpha_i \in L_i$ be such that $T_i' = Cn(\alpha_i)$.*

Revision is then carried out as follows: let $T_i'' = T_i *_i \alpha_i$ where $*_i$ is a local revision operator[15] for L_i. The revised B-structure $\mathcal{B} * \alpha = \{(L_1, T_1''), ..., (L_n, T_n'')\}$. When $L_\alpha \cap L_i = \phi, T_i * \alpha = T_i$, i.e., we can leave T_i unchanged, saving computational time. We define $\mathcal{B} \dot{+} \alpha$ analogously, except that we use the operator $\dot{+}$ on the various T_i. The i-shadow is the theory that represents the impact of the new formula in the agent's language L_i. When we revise by the i-shadow we do not expand the language of the theories in the B-structure. This means there can be loss of information.

Example: Our B-structure has theories $T_1 = Cn(p), T_2 = Cn(q)$. Then we receive the formula $\alpha = p \rightarrow \neg q$. But α cannot be split between L_1, L_2. The i-shadow of α turns out to be $\emptyset$ for both theories and no revision will take place. Here there is a loss of information, since the new information has cognitive value but we cannot use it.

Example: Our B-structure has theories $T_1 = Cn(p), T_2 = Cn(q)$. Then we receive the formula $p \wedge \neg q$ whose i-shadow is p for L_1 and $\neg q$ for L_2. T_1 is unchanged and T_2 is replaced by $Cn(\neg q)$. All of the new information is absorbed.

While there is loss of information (as in the first example), this is a tradeoff that an agent accepts because of the cognitive benefits (resource wise) that accrue to the agent as a result of maintaining its beliefs in a B-structure.

We remark that k-partitions are preserved by Option **A** revision but that m-consistency is not necessarily preserved by Option **A** revision.

Example: Agent A with B-structure $L_1 = \{t, p, r, s\}, L_2 = \{r, t, q\}, T_1 = Cn(t, p, \neg s \vee \neg r), T_2 = Cn(t \vee r, q)$ (the B-structure is 2-consistent) receives a new information $\neg r$ and adds it to both theories, rendering the B-structure 1-consistent.

Theoretically, a B-structure could become less coherent over time as more information is received. However, if an agent consistently receives information from *reliable* sources, the degree of consistency could actually increase. Presumably college graduates are more consistent in their beliefs than are people who have only graduated high school!

7.1.1 Relevance Relations in Option A Revision

An alternative way to understand option **A** revision is by the notion of a relevance relation between a new input and a theory in a B-structure. The existence of such a relation is an indicator of whether

[14] The i-shadow of α is similar to $forget(\alpha, L - L_i)$ where the operator $forget$ is as defined in [23]. However, note that the i-shadow of α depends only on α and L_i and therefore, bringing in L and then forgetting $L - L_i$ seemed unintuitive to us.

[15] We allow ourselves the assumption of the existence of these local revision operators, taking the view that for the B-structure approach, it is the *choice* of theories to be revised that is interesting.

the new formula brings about an increase in the information associated with a theory. In the case of Option **A** revision some theories are left untouched since they are considered irrelevant and amongst the theories that are revised, the only ones that actually change are those to whom the new formula brought new information. We offer the following definition:

DEFINITION 14 *A formula α is* relevant *to a subtheory T_i, $\mathcal{R}(T_i, \alpha)$ iff:*
1. $L_i \cap L_\alpha \neq \emptyset$, and
2. $((T_i \dot{+} \alpha) \cap L_i)$ is a proper *extension of T_i i.e., $\mathcal{M}((T_i \dot{+} \alpha) \cap L_i) \subset \mathcal{M}(T_i)$.*

In B-structure revision, the relation $\mathcal{R}(T_i, \alpha)$ between a sub-theory T_i and a new belief α determines which theories should get revised upon receipt of a new belief:

OBSERVATION 3 *If $L_\alpha \cap L_j \neq \phi$ and if $\mathcal{R}(T_i, \alpha)$ then $T_i * \alpha \neq T_i$.*

That is, a new piece of information is relevant to a sub-theory if the addition of the new belief to the sub-theory results in a change in the sub-theory.

The relation $\mathcal{R}(T_i, \alpha)$ is illustrated by the following example. Take $T_1 = Cn(p \vee q)$, $\gamma_1 = \neg(q \vee r)$, $\gamma_2 = q \vee r$, $L_1 = (p, q)$ and $L_2 = (q, r)$ Here, $\mathcal{R}(T_1, \gamma_1)$ obtains, but, as may be verified, $\mathcal{R}(T_1, \gamma_2)$ does not. An agent with the theory T_1 will not change its beliefs on receiving γ_2 since $T_1 * \gamma_2 = T_1$, but will do so on receiving γ_1.

7.2 Option B - The Merging Option

Given a formula α as input and a belief structure $\mathcal{B}$ we carry out Option **B** revision by a merger of subtheories that are affected by new information. Let $\Gamma_\alpha = \bigcup\{\Gamma_i : L_i \cap L_\alpha \neq \emptyset\}$, where L_α is the smallest language of α. Let $T_\alpha = Cn(\Gamma_\alpha)$. Now replace all languages L_i such that $L_i \cap L_\alpha \neq \emptyset$ by the single language $\bigcup\{L_i : L_i \cap L_\alpha \neq \emptyset\}$, which is their union. At the same time, replace all the corresponding T_i by the theory $T_\alpha * \alpha$.

This option will specialize to the procedure in [29] where the languages were all assumed to be disjoint but the receipt of information resulted in joining those theories whose languages overlapped the language of the new information. Option **B** style revision might not be the most plausible way to model revision as a general method since every new belief received leads to a merger of sub-theories (leading in turn to the progressive 'lumping' together of a B-structure over time). It is, however, a plausible method in some contexts. Generally, Option **B** results in two formally distinct subject areas merging as the result of some new information which straddles them (in real life this is likely to happen only occasionally).

Examples:

- Suppose I have a B-structure which keeps my beliefs about Turkey, Iraq and Iran separate. If I now receive a great many pieces of information about the Kurds who are scattered over these three countries then I may simply create the new subject *Asia Minor* and give up my attempt to deal with the three countries separately.

- A graduate student starting work on his dissertation maintains separate beliefs about the field of belief revision and non-monotonic reasoning. The student then reads an article by David

Makinson that points out the connection and intertranslatability between the two fields. From that point onwards, the student maintains his beliefs about the two fields together. The two areas are no longer separate.

One argument in favor of Option **B** style revision is that it takes care of problems with using a non-adjunctive, compartmentalized logic such as ours. Tanaka notes the following:

> once a belief set is compartmentalized, there is no way to remove the compartments if necessary. For example, as a result of compartmentalization of a belief set, beliefs about bus timetables are divided from those about train timetables,...,then one cannot infer which combination of bus and train makes the interchange more efficient. If reorganization of the belief set is said to be the key to resolve the situation, it must be that belief revision is not formalizable by non- adjunctive systems[16].

With Option **B** style revision, however, it is easy to imagine how the two subtheories corresponding to bus and train timetables could be merged into a new theory. For example, the agent receives a new piece of information that connects the two: public transportation timetables for the city that the agent lives in.

First Order B-structures

The introduction of a language with quantifiers introduces complications. We content ourselves with a few preliminary remarks. The *i-shadow* in Option **A** requires that each of the theories be finitely axiomatizable. This will be the case when we are dealing with a finite propositional language, but when we consider first-order languages (or others) we might not be able to make such an assumption. The theory however is still axiomatizable by some countably infinite, r.e. set of sentences. So, using revision operators for a possibly infinite set of sentences could be a technical solution for this problem.

In the propositional case we also made use of the following fact. Let T, T' be propositional theories whose languages are disjoint. If T, T' are individually consistent, then so is $T \cup T'$. Indeed, if v, v' are truth assignments for the languages of T, T' which make T, T' respectively true, then the truth assignment v'' which agrees with v on the propositional variables of T and with v' on the propositional variables of T', is a model[17] of $T \cup T'$. Intuitively, T and T' do not affect each other.

This simple fact can fail in the first order case. For example, if T contains the formula $(\forall x, y)(x = y)$ and T' contains the formula $(\exists x, y)(x \neq y)$, then T, T' will conflict *even though* they are individually consistent and have disjoint languages.[18]

To achieve the desired property of propositional theories also in the first order case, we make use of Robinson's joint consistency theorem[5]. Let be two first order theories in languages L, L' respectively and let $L'' = L \cap L'$. If $T \cap L''$ and $T' \cap L''$ are equal and complete, then $T \cup T'$ is a consistent theory in $L \cup L'$. In the case we are interested, $L''$is the language of the equality symbol $=$ and it is easy to see that it is sufficient that $(T \cap L'')$ is consistent with $(T' \cap L'')$. With this condition imposed, the developments in the propositional case can be extended to the first case. However, we do not have the space here to carry out a detailed development.

[16][34], page 35.

[17]Thus if T is true under 2 truth assignments and T' is true under 3, then $T \cup T'$ will be true under 6, which are obtained by combining any of the first 2 with any of the last 3.

[18]Equality is considered to be a logical symbol and not *part of the language* like a constant c or a unary predicate symbol M. So having the equality sign in common does not imply having common language.

8 Rules of Inference for B-structures

What is the logic of the agent whose beliefs are represented by a B-structure? What sort of inference operation have we defined on B-structures? How far from a classical propositional calculus does this agent deviate?

We define inference (i.e., we define a new inference operator $|\!\vdash$) from B-structures as follows:

DEFINITION 15 *Given a B-structure $\mathcal{B}$ and a formula γ, $\mathcal{B} \,|\!\vdash \gamma$ iff $v_{\mathcal{B}}(\gamma) \geq_k$ true (Cf. definition 10).*

Now, given a B-structure $\mathcal{B}$, formulas α, β, γ the questions below can be asked with respect to the status of the classical rules of inference. Here, $\mathcal{B} * \alpha$ is understood to be Option **B** revision (or Option **A** revision with the condition that any new epistemic input α (or β below) is in some L_i or, at least that the new input α (or β) $\Leftrightarrow \delta_1 \wedge ... \wedge \delta_p$ where each δ_k is in some L_i and p is small).

- **Disjunction Introduction:** If $\mathcal{B} \,|\!\vdash \alpha$ then does $\mathcal{B} \,|\!\vdash \alpha \vee \beta$?
- **Deduction:** If $\mathcal{B} * \alpha \,|\!\vdash \beta$ then does $\mathcal{B} \,|\!\vdash \alpha \rightarrow \beta$?
- **Conjunction Elimination:** If $\mathcal{B} \,|\!\vdash \alpha \wedge \beta$ then does $\mathcal{B} \,|\!\vdash \alpha$?
- **Disjunctive Syllogism:** If $\mathcal{B} \,|\!\vdash \alpha \vee \beta$, and $\mathcal{B}*\alpha \,|\!\vdash \gamma$, $\mathcal{B}*\beta \,|\!\vdash \gamma$, then does $\mathcal{B} \,|\!\vdash \gamma$?
- **Modus Ponens:** If $\mathcal{B} \,|\!\vdash \alpha \rightarrow \beta$ and $\mathcal{B} \,|\!\vdash \alpha$ then does $\mathcal{B} \,|\!\vdash \beta$?

The status of the rules above is provided by the following theorem:

THEOREM 8.1 *Assume that α, β, γ are such that $L_\alpha = L_\beta = L_\gamma$ where γ is the relevant compound formula formed from α, β and logical connectives. Then the inference operator $|\!\vdash$ for a B-structure $\mathcal{B}$, satisfies all the rules above. Moreover, for the first two cases above, we can weaken the condition to $L_\alpha, L_\beta \subseteq L_\gamma$.*

Proof: The result above is an immediate consequence of the more general fact that if $L_\alpha \subseteq L_\beta$, $\alpha \vdash \beta$ (where $\vdash$ is classical entailment), and $\mathcal{B} \,|\!\vdash \alpha$, then $\mathcal{B} \,|\!\vdash \beta$.

This result shows that classical rules for inference obtain in inference from B-structures when the agent restricts itself to a particular subject matter; once subject matters overlap, classical rules for inference can fail due to the interaction of beliefs about different subject matters conflicting with one another. Conversely, the plausibility of the method of query answering and inference defined on B-structures is brought out in cases where an agent might not have enough information to answer a particular query since only a small part of its B-structure has been activated. Once more information is supplied by the questioner, the agent can provide definite answers.

9 Computational Complexity

We now summarize the computational costs of B-structure query answering and revision. In general, both options **A** and **B** are *computationally efficient.* We assume that each L_i has relatively small size, say under some fixed p, while the cardinality N of the whole language $L = \bigcup L_i$ might be quite large ($\forall i, |L_i| \ll |L|$). Given a k-partition and a formula α with at most l distinct symbols, the query answering procedures run in time which is exponential in $l \times k \times p$, but *linear* in N. For option **A**, the procedure is linear in N, k and exponential only in $l \times p$. Thus if $k \times l \times p$ – the number of atomic propositions *relevant* to α – is small compared to N, as is usually the case, the computational cost will be much smaller than that of usual update procedures which are exponential in N. The complexity of belief revision procedures for B-structures compares favorably, then, to those of standard belief revision procedures. Whereas the original parameter size for the revision was N, the new parameter size is $k \times l \times p$ where $n, k, l \ll N$. This means that in the first case, we have a search space of size 2^N and in the second case we have a search space of size $2^{k \times l \times p}$. Such a reduction indicates considerable savings on computational costs.

Herzig and Rifi [20] have shown that the problem of calculating the smallest language for a formula is in co-**NP** in the size of the formula. However, since we think of an existing belief structure as being large and any particular incoming formula α as being relatively short, a problem which is **NP**-complete in the length of α will still be tractable. As for the formulas in the belief structure, they also should be put in their smallest language, but this is a one-time procedure for each individual formula and would not be expensive. In other words, a problem which is admittedly exponential in each individual formula, but only linear in the *number* of formulas in the belief base should not be too painful to deal with.

10 Review of Previous Work

We briefly discuss how other authors have addressed the issues raised in our study. Work on minimal change in revision has concentrated on minimizing the number of beliefs given up during the *contraction* operation. To this end, operations such as partial meet contraction using selection functions were defined in Alchourron *et. al.* [2]. A motivation for the choice of beliefs to be dropped is given by the notion of *epistemic entrenchment* introduced by Gärdenfors and Makinson [15] and refined for iterated belief change by Nayak [26] and Darwiche and Pearl [9]. Williams [36] uses the concept of *maxi-adjustment* to achieve maximal inertia of information under iterated belief revision. The notion of *epistemic relevance* is used for minimal contraction in Hansson [18] and Nebel [27]. Hansson and Wasserman ([19], [35]) have developed models for local change and inconsistency tolerance based on relevance relations between formulas in the belief representation.

The distinction between implicit and explicit beliefs, has been explored by the proponents of the *belief base* method such as Fuhrmann [14], Nebel [27], Hansson [18, 17]. Rott [33] combines some intuitions in showing how epistemic entrenchment orderings can be carried out for *safe contractions* for belief bases.

Belief revision for inconsistent belief bases has been studied in an alternative approach by Brewka [4]; paraconsistent belief revision is explored by Tanaka [34] while Restall and Slaney [31] have developed a paraconsistent semantical representation based on the revision of models approach suggested in Grove [16]. The investigation of complexity procedures in Nebel [27], via a fine-grained set of complexity classes, has shown that the complexity of base revision procedures satisfying AGM postulates is that of ordinary propositional derivability. Nebel's comparison of different revision methods shows that model-based revision methods such as those of Dalal [7] have a complexity which exceeds **NP** and **co-NP**.

11 Conclusion

We started this study by indicating the desiderata that a model for belief revision should meet. The B-structures framework meets all such desiderata. The B-structure model's accommodation of inconsistency is plausible and satisfying for several reasons. It provides a richer structure that explains *why* inconsistencies might be held *and* tolerated by the concept of overlapping subject matters. Such a notion is a plausible psychological description of human agents' reasoning.

In this model, minimal change is taken care of by its relevance sensitivity. A new epistemic input can only conflict with those parts of the B-structure that contain beliefs about the same (or related) subject matters. This protects the remaining parts of the B-structure from being affected by new inconsistencies. There is very little chance of a trivial update at the global level in a B-structure even though such changes might happen at the local level i.e., individual T_i in a B-structure could still be subject to the trivial update operation.

The B-structure's companion notion of inference enables a distinction between the explicit and implicit beliefs of the agent and shows how an agent with inconsistencies present in its belief representation can still be coherent. The logic of this inference is non-standard: several classical rules of inference of the propositional calculus lose their universal applicability and instead become language-sensitive. Since we have represented the subject matter of a proposition by its smallest language, this builds an interesting relevance sensitivity into the logic of the agent.

The issue of refinement raises a related issue, that of periodic splitting of B-structures given new contexts. One assumption of the B-structure model is that we are given a certain B-structure as the object of revision whose coherence could change over a period of time. Therefore, one avenue of investigation in the B-structure model is to devise a re-splitting operator, one that corresponds to an agent settling upon a particular, new, re-organization of its beliefs that suits its new cognitive objectives. Such a resplitting could also be undertaken by the agent after a review of its beliefs that are more exhaustive than those it normally permits itself. This could correspond to a second-order check on the accuracy and truth of its beliefs. The result of such an investigation could be a re-organization of its belief structures. The formalization of such a notion must wait for the time being.

Since a B-structure is a set of theories, the problem of finite representations is not taken care of. A possible alternative representation might be a B-structure that is a set of bases–yet another issue that we leave for future exploration. Having presented this framework, it is natural to ask, *what sort of agent have we modeled*? Our agent has beliefs about lots of different subjects, is possibly inconsistent when the totality of its beliefs is considered, does not like to think about more than what is absolutely necessary for a given context, but will combine information from related subject matters when answering a question or making its beliefs known, and when confronted with a inconsistency, is often willing to put up with it. It is our hope that the model meshes well with the intuitions that we have about how agents reason and provides a plausible, tractable method for belief representation and answering queries. Future work on this model will include the implementation of the query answering and revision procedures.

Future Work

There are two issues which we have not touched upon to any depth. One is the existence of a model theory and some sort of completeness result for our formal framework. Without such, how can we know what value our model has in purely logical terms? Of course we have shown some nice properties to hold, but what we would like to prove (and we believe it exists) is a maximality result. A completeness result for a formal system which is sound shows that every valid formula is provable. A consequence is that among sound systems, a complete one is *maximal*. We are dealing with frameworks which tolerate inconsistency and we cannot prove soundness in the usual sense. But we could hope to show that among desirable systems with certain features, ours is maximal so that a system with a more generous notion of inference would run into problems.

A second issue is pragmatic. Since we are dealing with inconsistent systems we can hardly claim that all the consequences will be *true*. However, [28] shows that an inconsistent extension of Peano arithmetic proves only true formulas provided that we concentrate on formulas which are *solely about numbers* and whose proofs are short. It should happen also with B-structures used in practice that in some sense they usually *work*. The wily airline stays in business and the bank continues to lend money for mortgages. So there should be a criterion for a B-structure to be pragmatic.

Dealing with these two issues would go far to substantiate our claim that B-structures are a new and important way of looking at reasoning. That work remains for the future.

References

[1] Special issue on relevance. In *Artificial Intelligence*, volume 97(1-2). 1997.

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

[3] N.D. Belnap. A useful four-valued logic. In J.M Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*. D. Reidel.

[4] Gerhard Brewka. Belief revision in a framework for default reasoning. In Andre Fuhrmann and N. Morreau, editors, *The Logic of Theory Change*, Lecture Notes in Artificial Intelligence, Number 465, pages 206–222. Springer, Berlin, 1991.

[5] C.C Chang and Jerome Kiesler. *Model Theory*. North-Holland, 1990.

[6] Samir Chopra and Rohit Parikh. An inconsistency tolerant model for belief representation and belief revision. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pages 192–197. Morgan-Kaufmann, 1999.

[7] Mukesh Dalal. Investigations into a theory of knowledge base revision. In *Proceedings of the 7th National Conference of the American Association for Artificial Intelligence*, pages 475–479, Saint Paul, Minnesota, 1988.

[8] Adnan Darwiche. A logical notion of conditional independence: properties and applications. *Artificial Intelligence*, 97(1-2)(8):45–82, 1997.

[9] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. In *Proceedings of Theoretical Aspects of Rationality and Knowledge*, pages 5–23, 1994.

[10] Richard Epstein. *The Semantical Foundations of Logic: Propositional Logics*. Oxford University Press, 1995.

[11] Melvin Fitting. Billatices and the theory of truth. *Journal of Philosophical Logic*, 18:225–256, 1989.

[12] Melvin Fitting. The family of stable models. *The Journal of Logic Programming*, 17:197–225, 1993.

[13] Melvin Fitting. A theory of truth that prefers falsehood. *Journal of Philosophical Logic*, 26:477–500, 1997.

[14] Andre Fuhrmann. Theory contraction through base revision. *Journal of Philosophical Logic*, 20:175–203, 1991.

[15] Peter Gärdenfors and David Makinson. Revisions of knowledge systems using epistemic entrenchment. In Moshe Vardi, editor, *Proceedings of Theoretical Aspects of Reasoning about Knowledge*, pages 83–96, 661–672. Morgan Kaufmann, 1988.

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

[17] Sven Ove Hansson. A dyadic representation of belief. In Peter Gärdenfors, editor, *Belief Revision*, Theoretical Computer Science. Cambridge University Press, Cambridge, 1992.

[18] Sven Ove Hansson. In defense of base contraction. *Synthese*, 91:239–245, 1992.

[19] Sven Ove Hansson and Renata Wasserman. Local change. Technical Report PP-1999-17, Institute for Logic, Language and Computation, 1999.

[20] Andreas Herzig and Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence*, 115(1):107–138, 1999.

[21] Henry Kyburg. *Probability and the Logic of Rational Belief*. Middletown, Wesleyan, CT, 1961.

[22] Henry Kyburg. *Epistemology and Inference*. University of Minnesota Press, Minneapolis, MN, 1983.

[23] Jerome Lang and Pierre Marquis. Complexity results for independence and definability in propositional logics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference*, pages 356–367, San Mateo, CA, 1998. Morgan-Kaufmann.

[24] D. McDermott. We've been framed, or why ai is innocent of the frame problem. In Z. Pylyshyn, editor, *The Robot's Dilemma*, pages 113–122. Ablex, 1987.

[25] Marvin Minsky. *The Society of Mind*. Simon and Schuster, 1986.

[26] Abhaya Nayak. Iterated belief change based on epistemic entrenchment. *Erkentennis*, 41:353–390, 1994.

[27] Bernhard Nebel. Syntax based approaches to belief revision. In Peter Gärdenfors, editor, *Belief Revision*, Theoretical Computer Science. Cambridge University Press, Cambridge, 1992.

[28] Rohit Parikh. Existence and feasibility in arithmetic. *Journal of Symbolic Logic*, 1971.

[29] Rohit Parikh. Beliefs, belief revision, and splitting languages. In Lawrence Moss, Jonathan Ginzburg, and Maarten de Rijke, editors, *Logic, Language, and Computation, Volume 2*, CSLI Lecture Notes No. 96, pages 266–268. CSLI Publications, 1999. Initially presented in *Preliminary Proceedings of Information Theoretic Approaches to Logic, Language, and Computation 1996*.

[30] Pavel Pudlak and Sam Buss. How to lie without being easily convicted and the length of proofs in propositional calculus. In Pacholski and Tiuryn, editors, *8th workshop on Computer Science Logic, Kazimierz, Poland, September 1994*, Lecture Notes in Computer Science # 933. Springer Verlag, 1995.

[31] Greg Restall and J. Slaney. Realistic belief revision. Technical Report TR-ARP-2-95 Automated Reasoning Project, Australian National University, 1995.

[32] Odinaldo T. Rodrigues. *A methodology for iterated information change.* PhD thesis, Imperial College, University of London, 1997.

[33] Hans Rott. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information*, 1:45–78, 1992.

[34] Koji Tanaka. What does paraconsistency do? the case of belief revision. *The Logica Yearbook*, pages 188–197, 1997.

[35] Renata Wassermann. On structured belief bases. In Hans Rott and Mary-Anne Williams, editors, *Frontiers in Belief Revision.* Kluwer, 1999. To appear.

[36] Mary-Anne Williams. A practical approach to belief revision: Reason-based change. In L. Aiello and S. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference*, pages 412–421, San Mateo, CA, 1996. Morgan-Kaufmann.