ABSTRACT

Title of dissertation:	TOPIC-SENSITIVE BELIEF REVISION
	Bijan Parsia, Doctor of Philosophy, 2009
Dissertation directed by:	Professor John F. Horty Department of Philosophy

When asked to change one's beliefs in the face of new information, or to revise a book given errata, we commonly strive to keep our changes relevant, that is, we try to restrict the beliefs (or chapters) we change to those that bear some content relation to the new information. One kind of relevance, topicality, is interesting for two reasons: First, topicality tends to be strongly encapsulating, e.g., we shouldn't make any off-topic changes. Second, topicality tends to be weaker than strict relevance. Consider a panel of three papers on the topic of Kant's life and works. It would be entirely possible for each of the papers to have no bearing on the truth of any sentence in any of the other papers, and yet for all of the papers to be on topic.

In this dissertation, I explore theories of logical topicality and their effect on formal theories of belief revision. Formal theories of belief revision (in the Alchourròn, Gärdenfors, and Makinson (AGM) tradition) model the object of change (my beliefs, a book) as a collection of formulae in a supra-classical logic and provide a set of postulates which express constraints on the sorts of change that are, in principle, formally rational. In 1999, Rohit Parikh proposed that signature disjointness captured a reasonable notion of topicality but that taking topicality into account required changes in the standard AGM postulates (and thus, the notion of rational change). He, and subsequent theorists, abandoned this notion of topicality in order to deal with the revision of inconsistent objects of change. In this thesis, I show 1) that a disjoint signature account of topicality does not require changes to the AGM rationality postulates and 2) a disjoint signature account of topicality can apply to inconsistent objects of change. Additionally, I argue that signature disjointness has a strong claim to being at least a sufficient condition of logical topicality.

TOPIC-SENSITIVE BELIEF REVISION

by

Bijan Parsia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Commmittee:
Professor John F. Horty, Chair/Advisor
Professor Peter Carruthers
Professor Mathias Frisch
Professor Paul M. Pietroski
Professor V.S. Subrahmanian

Acknowledgements

I'd like to thank my beloved, Zoe Mulford, for writing so many kick ass songs and for putting up with the severe snuggle deficit that was the direct consequence of the late nights finishing this thing. Also, for proofreading it. It's a comfort to know that someone else has read every word.

If I didn't thank my mother, Maria Leonard, then the gender balance of this acknowledgements would be off. So, thanks mom! Big thanks for dumping my father!

Jeff Horty made this redemption possible and put up with my recalcitrance with graceful understanding.

Uli Sattler was a nag and a help and a helpful nag.

Birte Glimm rocked on the moral support.

So did Yarden Katz before MIT swallowed him up.

Jim Hendler hired me on the basis of a throwaway IRC comment and thus brought me back into academic life.

But of all things in this world, I think I must acknowledge that I probably would have taken several more *years* if not for 20mg of Citalopren, once a day. Thanks lil pill! You worked wonders. And thanks to Dr. Chris Dickens for prescribing it.

Table of Contents

Chapter 1

Introduction

One of the strong appeals of regimenting theories within a formal logic is the clarity of the resultant regimented theory. The logical commitments of the regimented theory are exactly the logical commitments of the axiomatization itself. One is forced to be precise where vagueness used to reign, and complete where hand-waving once sufficed. (The adequacy of the regimentation as a presentation of the original theory is a different story.)

With the advent of sufficient computation power to make state of the art theorem provers practical for a large variety of axiom sets, regimented theories become even more appealing. Not only can the formalized theories be used in programs for a variety of purposes, but automation can make large-scale theory construction feasible. Syntactic correctness, consistency, and verified consequences are all crucial to developing large theories, but there are an increasing number of services that can be mobilized, from the quotidian (version control; document management) to the more exotic (concept unification and other search techniques; visualization; explanation). The field of knowledge representation (KR) and reasoning (&R) in part focuses on developing formalisms that, while representationally adequate for a domain, lend themselves to effective automation.

Large machine-oriented formal theories in expressive logics (e.g., those with disjunction and classical negation) present a number of challenges. Their sheer size in conjunction with the intractability of the logic can overwhelm systems. Similarly, even with good search and visualization capabilities, human users can easily get lost. This tendency can be exacerbated when the theory concerns a number of diverse subject matters. Large, shared, evolving theories often do, especially if they incorporate application-dependent extensions to theories developed by other people.

Reuse of theories by sheer aggregation is not a generally workable strategy for several reasons. It is easy for aggregated theories to be too large both for people and programs to effectively handle. Uncoordinated development of theories can contain contradictions and other infelicities which require attention to the whole of the unwieldy aggregate, as well as the source theories. Finally, aggregation, absent very careful factoring, brings in too much: the relevant and the irrelevant; the center, the periphery, and the rather random digressions.

People use all sorts of filters to manage the array of sentences they are nominally committed to. They give differential weight to trusted proponents and schools of thought. They are inattentive to consequences and connections. They also compartmentalize based on context and concern. In particular, for particular tasks in particular circumstances, they identify *topics* of concern, and the subset of sentences that are relevant to that topic. These might be disconnected and confused; contradictory; or coherent and comprehensive. In other words, there might be a subset of the sentences I am generally prepared to mobilize that deal primarily with the topic at hand and deal with it well. Furthermore, I am certainly not compelled, by the dictates of rationality, to root out all the difficulties of the broader set of sentences before pronouncing on a topic — even, or perhaps especially, with authority. Expertise critically depends on being able to focus on the subject of discussion and is not directly damaged by kookiness in other areas. Challenges based on the strangeness of irrelevant beliefs are *ad hominem*, almost by definition (such challenges are paradigmatic of the fallacy). However, they may be non-fallaciously mobilized to shift the burden of proof, for example. If one's general competency is impugned by some beliefs, it is reasonable to raise those beliefs regardless of topic.

Given a diverse, formalized theory captured as a knowledge representation[1], it is reasonable to wonder if the strategies used by human cognizers have a place. Part of the point of mechanization is to transcend human limitations such as that of attention span, so one might think that human focusing and scoping techniques are not of interest for knowledge representation. After all, theorem provers can detect contradictions in knowledge bases far too large for any human to survey, much less comprehend without the need for significant chunking of the representation. Of course, on the face of it, such wondering is a bit silly as knowledge representations are targeted at people as well as programs:

> Finally, knowledge representations are also the means by which we express things about the world, the medium of expression and communication in which we tell the machine (and perhaps one another) about the world. This role for representations is inevitable so long as we need to tell the machine (or other people) about the world, and so long as we do so by creating and communicating representations. The fifth role for knowledge representations is thus as a medium of expression and communication for use by *us*.

[1] I shall use the term "knowledge representation" primarily to talk of computational *artifacts*, that is, of logical theories with a concrete encoding in a language meant for mechanical processing, especially automated reasoning.

[Davis et al., 1993]

However, the worry can be sharpened. Given the support of automated reasoning (and similar representation processing tools), are the kinds of filtering and focusing people do worth emulating? Can other services do better? As I mentioned above, size can be a problem in and of itself. Even if the logic underlying a knowledge representation has decidable inference problems, it is rare that those will be (worst case) tractable. After all, even propositional logic satisfiability is NP-complete. So, even small representations can be hopeless if they hit the worst case, or simply if they are not amenable to state of the art optimizations. Similarly, if the representation has logical difficulties such as contradictions, they must be managed somehow. For contradictions, either they must be resolved by people (with the aid of "debuggers"), resolved entirely by a program, or the logic must be paraconsistent. In the first case, the problem of attention remains even with significant automation, and worse, the problem might lie in a portion of the representation outside the expertise of the human trying to resolve the contradiction. In the second case, even if computationally feasible, there are serious issues of non-determinism and trust. Users will typically want to understand the repair performed (and the ramifications thereof), at least in general terms, and also need good reason to accept the resultant situation. Since no one can hold very large theories in mind (or even sensibly work with them manually), some sort of filtering will be required just to understand repairs, even if it is just listing the specific axioms changed. While many changes do not have "global" effect, some do — the obvious case being the introduction of a contradiction. To present a change as a reasonable repair, the system needs to show what parts of the representation are significantly affected. More importantly, the affected parts have to *matter* to the user and their current application. In the third case, the logic must be (in broad terms) comprehensible and, in general, appropriate. While all paraconsistent logics avoid, at least in some cases, *ex contradictione quodlibet*, they do so in significantly different ways and answer to a wide range of logical intuitions.

For example, non-adjunctive logics (i.e., logics without or with a weakened &-introduction) can be plausibly read (following [Jaśkowski, 1969]) as modeling the aggregation of belief sets where each belief set must be consistent. That is, contradictions *within* a belief set are treated as devastating, where contradictions *between* belief sets are handled more gently, which, given the prevalence of disagreement, is wise. But there are many ways of being gentle, not equally pleasing, but also not clearly ranked. One could take the "consensus" view of the beliefs sets (i.e., the intersection of the belief sets), or a more expansive, "non-disputatious" view (i.e., where every belief without a contradictory belief in any belief set is accepted), or impose an ordering on the belief sets where higher ranked sets trump lower ranked one where there is a dispute and aggregates otherwise.

When the "belief sets" are in fact large multi-topic knowledge representations, then these three options generate either a minimal (consensus) or nigh-maximal (non-disputatious) or maximal common, consistent (presuming each is self-consistent) representation. But are any of these desirable? Certainly from the perspective of having a consistent representation that contains some or most of the information in the original representations, all three possibilities, properly conceived, deliver the goods. However, consider the following three representations:

$$B_1 = \forall.x(P(x,y) \rightarrow C(y)$$

$$B_2 = \forall .x(P(x,y) \rightarrow C(y)$$

$$B_3 := P(a,b)$$

On the non-disputatious view, considering only the explicit formulae, the result is the union of B_1, B_2, and B_3. However, their union is inconsistent (since b is forced to be both C and $\neg C$). Even if we consider non-disputatiously unioning their respective deductive closures, we get an inconsistent set.

If one is only concerned with a particular topic, dealing with the whole of each representation as such is wasteful at best unless they each are about that topic alone. Intuitively, there is no need to reconcile parts of representations that do not bear upon the current purpose. One may wish to defer such reconciliation for as long as possible if only to avoid work if those parts are never used. Also, there is a discomfort in meddling with parts of a representation which are not pertinent. It seems risky. Given that in many logics axioms can have highly non-local effects — even discounting the effects of contradictions — it is sensible to be conservative in the adaptations one makes. Of course, if one is concerned with the amalgamated representation itself, then its entire rectification is of concern (though, in a practical setting, one might chose to focus on certain bits at a time, or to divide the labor between various experts). If one is merely trying to reuse the representation in a particular context for a particular purpose, then there are considerations driving a tighter focus. These task-specific considerations might conflict with more general ones. In the above example, one might have a general preference for one belief set over another based on the general reliability of the advocates for each. However, the generally less reliable might be idiosyncratically highly reliable on certain matters. Context sets certain parameters for the evaluation of repairs to the amalgam, and brings some criteria of relevance.

Again, considering only the problem of contradictions, we are led to consider relevant logics. Presuming that the representation will be used as a data source to evaluate queries against (given a suitable proof theory), then various relevant logics will, for each query, be sensitive to only the relevant (given the signature of the queries) sentences in the representation:

> One of the fundamental ideas in relevance logic is to keep track of which sentences are used in the derivation of other sentences. If a derivation of sentence C is possible given sentence A, it does not follow that A relevantly implies C, for A may not have played any role whatsoever in the proof of C. One of the signs that a formula has played a role in the proof of another is if they contain common notation. In many relevance logics, A implies C only if there is a variable common to both A and C. This property allows the data that are relevant for proving a given conclusion to be neatly circumscribed. pg. 213[Garson, 1989]

Garson appeals to relevance logic and its circumscribing properties to characterize the partitioning of a (first-order) logical theory into what he calls *logical modules*. His purpose is to increase the performance of reasoning by dividing the task among domain-focused reasoners (perhaps with domain-tuned reasoning procedures), inspired by [Fodor, 1983]'s vision of cognition as a collection of specialized modules. A logical module has a number of features:

> A sentence is *correct* (for [some global logic,] G) just in case it is provable

> in G from the global data. A sentence is *provable in a module* just in case there is a proof of that sentence from the module's data, using the module's inference rules. A module is *locally consistent* for the global logic G just in case any sentence provable in a module is correct, and a module is *locally complete* (for G) just in case every sentence in the module's domain that is correct is provable in the module. A module is *logical* (for G) just in case it is both locally consistent and locally complete. A sentence is *provable in a reasoner* just in case the sentence can be proven in the module that is selected for it. pg. 210, [Garson, 1989]

The measure of adequacy of a logical module (and its associated reasoner) is the global logic, G. For the purpose of reasoning, we want modular fragments of a knowledge representation to *encapsulate* some aspect of the whole, in particular, enough so that we can perform the reasoning task at hand. Ideally, we would be able to reason with the modular fragment *as if* it were the whole. One could hope that this would improve efficiency of reasoning, either by avoiding unnecessary work by permitting tuned subreasoners (e.g., [Amir and McIlraith, 2005]), or by allowing parallel reasoning.

However, modularization for reasoner performance is not, in itself, an especially compelling application of modularization. From an automated reasoning perspective, enormous strides have been made without modular segmentation of representations expressed in all sorts of logics, from propositional to full first-order logic. Even in program optimization it is well known that whole program analysis [Allen, 1974] can dramatically improve the performance of programs.

In any case, when exploited for performance, modularity should be *semantically* neutral. That is, qua optimization, exploiting modularity should not affect the meaning of a representation (which is a problem for Garson's account since he requires shifting from first-order logic to a relevance logic).

For other tasks, modularity might play a more significant role. In particular it seems possible that the effect of *changes* to a representation might vary depending on whether one whises to respect the modular structure. As a limit case, one could imagine that the only difference between modular and non-modular theory change would be performance. That is, one could impose a constraint that the two forms of alteration should give the same results. However, it is not at all clear that this can be, or even should be, the case.

The purpose of this thesis is to investigate modular, specifically *topic oriented*, revision of knowledge representations.

Chapter 2

Background & Preliminaries

In this chapter we present a brief overview of the basics of computationally oriented belief revision.

2.1 Representations

While very little of the discussion in this thesis is *specific* to the revision of knowledge representations, it is strongly informed by such considerations. AGM style belief revision theory is, by design, very general and intended to be a suitable framework for informing investigations of change to psychological states, legal frameworks, books, or group consensus. However, the particular application in mind certainly affects the flavor of the discussion and the plausibility of various arguments. Ultimately, the results of this thesis will be applied to working with knowledge representations (specifically, those expressed in various description logics, i.e., fragments of first order logic). Thus, it behooves us to consider knowledge representation formalisms.

2.1.1 Logic

While there have been many proposals for representing knowledge including semantic nets [Collins and Quillian, 1969], frames systems [Minsky, 1974], programs and procedures [McDermott, 1987], production rules [Buchanan and Shortliffe, 1985], and much more exotic mechanisms [Gärdenfors, 2000], by far the dominant approach has been some sort of logic — typically a variant, subset, or extension of first order logic.[1] First order logic has a lot of attractive properties for KR including being exceedingly well understood ([McCarthy, 1958]). It is the standard example of a formalism with a model theoretic semantics and various sound and complete inference systems.

It's hard to overestimate the importance of these features. In the early history of KR, there were many proposals for novel mechanisms for encoding representations. However, these attempts were plagued by ambiguity, unclarity, and implementation dependence. Ron Brachman, famously in [Brachman, 1983], identified 29 different senses of "is-a" common in KR systems of the time, often without any explicit indication of which sense was in force in any particular representation at any particular point.[2] Frame systems and semantic nets have, at least in their monotonic aspects, been replaced with description logics [Levesque and Brachman, 1987, Hayes, 1979], while the non-monotonic aspects have been dealt with by a variety of non-monotonic logics including circumscription [Brewka and Augustin, 1987], logic programming [Kifer et al., 1995], default logic

[1]The three major extensions (or alternatives) are modal logic, higher order logic, and various flavors of non-monotonic logic. Of the three, only the last tends to break key assumptions of belief revision theory.

[2]For example, [Kashyap and Borgida, 2003] describes an attempt to migrate the "Semantic Network" (*SN*) portion (i.e., the knowledge portion) of the Unified Medical Language System to a description logic based representation: "*SN* types, relationships and their hierarchies, as well as inverses have clear corresponding OWL/DL constructs. However, there are serious difficulties in accurately capturing the semantics of the *SN*, due both to the under-specified meaning of the notion of "link" between two semantic types, and the somewhat unusual inferences/constraints that are associated with them in *SN*. " See [Woods, 1975] for a similar critique.

[Baader and Hollunder, 1995] (but see [Horty, 1994] for a counter both to a logic translation account and for issues with using default logic), or other exotica [Thomason and Horty, 1989].

For our purposes, it is sufficient to note that, following [Gärdenfors, 1988], we only consider logics which are propositionally complete, that is, contain propositional logic. We use the following connectives $\vee$, $\wedge$, $\neg$, $\rightarrow$, and $\leftrightarrow$ with their conventional truth functional readings of disjunction, conjunction, negation, material conditional, and material biconditional. As common, we use uppercase letters, $A \ldots Z$ for atomic sentences. $\models$ indicates entailment and we shall generally allow both single sentences and sets of sentences to appear on the left hand side as is convenient. We use $\mathrm{Cn}(X)$ operator to denote the deductive closure of the set of sentences X (i.e., $\mathrm{Cn}(X) =_{def} \{\alpha | X \models \alpha\}$). Additionally, $\mathrm{Sig}(S)$ denotes the set of non-logical vocabulary in the set or sentence S, i.e., the signature.

2.1.2 Extra-Logical Aspects

A distinctive feature of KR formalisms which distinguishes them from the logics that underlying them is that various non- or extra-logical aspects are critical aspects of the language. For example, the gensym fallacy notwithstanding [McDermott, 1976], the *names* used in a representation are often of critical importance to applications or to representation developers. For KRs which underlie controlled vocabularies, the *annotations* (i.e., textual or structural information associated with a term) are often the *point* of the representation, with the logical aspects serving as an organization tool [Sioutos et al., 2007]. Similarly, various sorts of metadata such as author, modifier, source, or date changed may have a great deal of influence on how the representation is used or evolved. In some cases, the underlying logic is made sensitive to such features (e.g., [Kifer and Subrahmanian, 1992]) but more often such features are outside the base formalism. In this thesis, we will ignore those features entirely, except to note that where there is a choice of change operator it may be that desirable operators are sensitive in specific ways to this extra-logical information.

2.2 The Standard Operations and their Postulates[3]

The starting place for any account of theory, knowledge, or belief change is the AGM[4] theory originally propounded in [Alchourrón et al., 1985] and influentially elaborated in [Gärdenfors, 1988]. The AGM account (and AGM-like accounts) comprises a model of the belief object, at least three transformation operators which go from the belief object to other belief objects, and a set of axioms which constrain the behavior of those operators. In standard AGM theory, the belief object is a *belief set*, that is, a set of propositions (formulae, sentences, what have you) which are closed under some (deductive) consequence relation. In other words, for any set of sentences, T, $Cn(T)$[5] is a belief set.

[3]The formulation of the postulates is taken primarily from Stanford Encyclopedia of Philosophy's article on belief revision, `http://plato.stanford.edu/entries/logic-belief-revision/`, and [Gärdenfors, 1988]. The names for the postulates are mostly taken from the former. The discussion, while standard, closely follows that in [Gärdenfors, 1988].

[4]Named after the prime movers, Alchourrón, Gärdenfors, and Makinson.

[5]One can parameterize the *Cn* operator by different consequence relations over (i.e., different semantics for) the same set of sentences (e.g., Datalog under minimal model vs. first order semantics with respect to existential queries [Horrocks et al., 2005]). In this thesis, the consequence relation is generally truth-

All of the operators take, as arguments, a belief set and an input (i.e., a new sentence) and return a modified belief set.

In AGM theory, the primary (at least, in the sense of *first*) presentation of the three operations is given as a set of postulates held to be true of any rational version of the operator. In the case of expansion, the postulates determine a univocal operational definition of the operator. However, for contraction and revision (which are interdefinable) this is not the case. The set of rational operators are (standardly) equivalently characterized in terms of a variety of other mechanisms (including operational). By their equivalents and the lack of any intuitive additional constraining abstract factors, the various accounts lend plausibility to each other.

2.2.0.1 Expansion (+)

Expansion is the operation of adding new beliefs to one's belief set. The addition is uncritical in the sense that nothing is removed in response to the input, even if the input contradicts the original belief set. Thus, expansion is simply the closure of the set theoretic union of the belief set and the input. Nevertheless, we can give a set of postulates that characterize this operator more abstractly:

- E1 Closure: $T + \alpha = \mathrm{Cn}(T + \alpha)$
- E2 Success: $\alpha \in T + \alpha$
- E3 Inclusion: $T \subseteq T + \alpha$
- E4 Vacuity: If $\alpha \in T$, then $T + \alpha = T$
- E5 Monotonicity: If $T \subseteq T'$, then $T + \alpha \subseteq T' + \alpha$
- E6 Minimality: $T + \alpha$ is the *smallest* belief set that satisfies E1-E5.

---

functional propositional logic.

If the input is *not* in conflict with the belief set, then expansion is sensible. When it *is* in conflict, i.e., when $\neg\alpha \in T$ and the input is α, then expansion trivializes the belief set by making it inconsistent. By Closure, expanding by a conflicting input adds every legal sentence of the language to the belief set.

Expansion to the entire language is not rational (not even in a paraconsistent view, since, after all, paraconsistent logicians argue that contradictions are *not* trivializing), thus expansion only makes sense when there is no conflict between the belief set and the input.

2.2.0.2 Contraction ($-$)

Contraction is a far more interesting operation on belief sets. Contraction is the operation of removing an input belief from one's belief set. It is thus the reciprocal (in some sense) of expansion. Unlike expansion, contraction is complex in several ways, but first let's examine the "core" postulates:

C1 Closure: $T - \alpha = \mathrm{Cn}(T - \alpha)$
C2 Inclusion: $T - \alpha \subseteq T$
C3 Vacuity: If $\alpha \notin \mathrm{Cn}(T)$, then $T - \alpha = T$
C4 Success: If $\alpha \notin \mathrm{Cn}(\emptyset)$, then $\alpha \notin \mathrm{Cn}(T - \alpha)$
C5 Recovery: $T \subseteq (T - \alpha) + \alpha$
C6 Extensionality: If $\alpha \leftrightarrow \beta \in \mathrm{Cn}(\emptyset)$, then $T - \alpha = T - \beta$

Closure, Inclusion, Vacuity, and Success straightforwardly parallel their expansion counterparts, at least in the formulation. If we consider Closure, it's clear that we cannot use set theoretic removal as a contraction operator. This fact is easy to illustrate:

Example 1. *Let $T = Cn(P \wedge Q)$. Consider $T' = T \setminus P$. Clearly $P \wedge Q \in T'$. Thus, when* *we take the Cn of T' to satisfy closure, we reintroduce P since $P \wedge Q \models P$. Thus, $\setminus$ is not an AGM rational contraction operator, as it violates Inclusion.*

As a consequence of the need to cope with entailment, not only can we not rely on simple set theoretic operators plus *Cn*, but there is no single, logically non-arbitrary operator that is determined by the postulates. This follows from the fact that there will often be multiple ways of breaking an entailment from a belief set. Thus, there is often the *possibility* of choice.

This fact is easier to see if we consider operationally based accounts (or, following, [Gärdenfors, 1988] "constructions") of contraction operators. The core tool for defining contraction operators is the notion of a maximal non-entailing subset, aka, a *remainder*:

Definition 1. *For a belief set, T, and a sentence, α, the* remainders *of T given α, $T \perp \alpha$ are all sets, t, such that:*

1. $t \subseteq T$
2. $t \not\models \alpha$
3. *there is no $t' \subseteq T$ such that $t \subset t'$ and $t' \not\models \alpha$.*

If we take the set of remainders for the negation of some input, i.e., $T \perp \alpha$, it is easy to see that the closure of any particular remainder will be a non-trivial belief set and, furthermore, than such an operator (i.e., one that creates a contracted belief set by picking the closure of one of the remainders as the output) will conform to the core postulates. Such operators are called *maxichoice* operators. They are "maximal" in exactly the sense that there are no more sentences in T that could be added to $T -_{maxi}$ and still have Success. Of course, since $T \perp \alpha$ can have a cardinality greater than one, merely being maxichoice is not sufficient to determine a unique operator.

The problem with maxichoice contraction is that the outputs are too large. Consider a simple example:

Example 2. *Let $T = Cn(P \wedge Q)$. Clearly, $P \vee Q$ is an element of T as is $P \vee \neg Q$. Now, consider $t \in T \bot P$. It is either the case that*

1. *$\{Q, P \vee Q\} \subseteq t$ and $P \vee \neg Q \notin t$ (since $\{P \vee \neg Q, Q\} \models P$), or*

2. *$\{P \vee \neg Q\} \subseteq t$ while $Q \notin t$ and $P \vee Q \notin t$ (since, additionally, $\{P \vee \neg Q \vee Q\} \models P$).*

So, every remainder contains either $P \vee Q$ or $P \vee \neg Q$ (though not both).

Given this face, if we *expand* by the negation of the contracted input (which defines revision, see section 2.2.0.3), we end up with a maximal belief set, that is, one where for every sentence the belief set contains either that sentence or its negation. This is a rather surprising and counterintuitive result, so maxichoice contraction is generally considered to be an inappropriate operator. Of course, one might instead take the *intersection* of all the remainders, that is, be more *cautious* in one's commitment and thus more ruthless in one's pruning. Such a contraction operator yields *full meet* contraction. Full meet contraction has the additional attractive property that it does determine a univocal operator—no choice involved. Unfortunately, full meet contraction is *too* ruthless and thus its results are too small:

Example 3. *Let $T = Cn(P \wedge Q)$. Clearly, $P \vee Q$ is an element of T as is $P \vee \neg Q$. Now, consider $t \in T \bot P$. It is either the case that*

1. *$\{Q, P \vee Q\} \subseteq t$ and $P \vee \neg Q \notin t$ (since $\{P \vee \neg Q, Q\} \models P$), or*

2. *$\{P \vee \neg Q\} \subseteq t$ while $Q \notin t$ and $P \vee Q \notin t$ (since, additionally, $\{P \vee \neg Q, P \vee Q\} \models P$).*

So, there is always at least one remainder which does not contain one of $\{Q, P \vee Q, P \vee \neg Q\}$. *Thus,* $T -_{full} P$ *will not contain any of those sentences, or, indeed, any other sentence that follows from* P *or from* Q. *In fact,* $T -_{full} P$ *will only contain sentences such as* $\neg P \vee Q$ *which follow from* $\neg P$ *alone.*

When combined with the expansion by $/P$ we end up with a revision that replaces the entire belief set with the input. While that may sometimes be desirable, it's clearly not a generally rational operator.

The middle ground operator, that is, that which takes a distinguished subset of the remainders and intersects them, yields *partial meet* contraction which avoids both the too much and the too little problem. One cost, as with maxichoice contraction, is that we do not have a purely logical determination of a rational contraction operation. Distinct partial meet contraction operators are generated by a variety of remainder selection mechanisms, most prominently by *priority* orderings on remainders. Such orderings are called *epistemic entrenchment* relations.[6]

If a selection function is defined in terms of an ordering over (all possible) remainders, then it is a *relational* selection function and generates a relational partial meet contraction operator. If the relation is transitive as well, then the generated operator is a *transitively relational* partial meet contraction operator. These two conditions on selection functions are significant not just for their naturalness, but because they correspond to

[6] Entrenchment can be defined over remainders or over sentences (for the latter, see 3.4 of [Gärdenfors, 1988]. While entrenchment orderings over sentences are more natural, they do not change anything essential about the operators. Since orderings on remainders are more convenient for the purposes of this thesis, we shall stick with them.

two additional postulates on contraction, the so-called supplementary postulates:

C7 Conjunctive Overlap: $(T - \alpha) \cap (T - \beta) \subseteq T - (\alpha \wedge \beta)$
C8 Conjunctive Inclusion: If $\alpha \notin T - (\alpha \wedge \beta)$, then $T - (\alpha \wedge \beta) \subseteq T - \alpha$

While generally less well regarded as the core postulates, both have some intuitive appeal. Conjunctive Overlap states that revising by both conjuncts of a conjunction is stronger than revising by the conjunction itself.[7] This seems fairly reasonable as the Success of $T - (\alpha \wedge \beta)$ is ensured by either $(T - \alpha)$ or $(T - \beta)$. Similarly, Conjunctive Inclusion observes that if contracting by a conjunction doesn't contain one of the conjuncts then revising by that conjunct is (equal to or) larger than revising by the conjunction.

What is striking, however, is that the core postulates plus Conjunctive Overlap describe exactly the set of relational partial meet operators, and the full set of postulates describe transitively relational partial meet operators. That is, if a contraction operator is a relational (resp. transitively relational) partial meet operator, then it satisfies C1-C7 (resp. C1-C8). Furthermore, any operator which satisfies C1-C7 (resp. C1-C8) is (or is equivalent to) a relational (resp. transitively relational) partial meet contraction operator. Since relational and transitively relational entrenchment are rather intuitive, they lend a great deal of plausibility to the corresponding postulates.

2.2.0.3 Interdefinability of contraction and revision

As discussed in the prior section, we can think of revision as a process of contraction then expansion. This reduction is called the *Levi identity*:

[7]Technically, the intersection is critical, since it does not give us an iterated contraction, such as $(T - \alpha) - \beta$. However, this does affect the key point.

Definition 2 (Levi Identity)**.** $T * \alpha =_{def} (T - \neg\alpha) + \alpha$

From the Levi identity we can show that if the plain expansion of target belief set (by a non-self-contradictory input) does not produce a contradictory belief set, then that revision reduces to plain expansion. This immediately follows from the Vacuity of contraction. Indeed, the postulates for revision, given the Levi identity, can be (mostly) straightforwardly derived from the postulates for contraction and expansion. These facts, and our current presentation suggest that contraction is somehow more fundamental than revision. (And, indeed, many authors take it to be so.) However, contraction is equally definable in terms of revision:

Definition 3 (Harper Identity)**.** $T - \alpha =_{def} (T * \neg\alpha) \cap T$

The Harper identity is less intuitive than the Levi identity (which seems to be a straightforward application of contraction and expansion!), but the idea is that revising by $\neg\alpha$ must remove everything that entails α (on pain of contradiction). If we then intersect the revision with the original belief set, we remove everything that entails $\neg\alpha$, which, of course, could not be in the original belief set (again on pain of contradiction, unless the original belief set did not contain α).

While the "direct" intuitions favor the Levi identity (and thus contraction as fundamental), we saw in section 2.2.0.2 that the problems with maxichoice and full meet contraction are much starker when considering their effects on revision. We might be able to fool ourselves into believing that the results of maxichoice or full meet contraction are sensible taken on their own terms, but certainly not when considering whether it is rational to always generate a maximal belief set or an obliterating belief set.

2.2.0.4 Revision ($*$)

With the connection to contraction well in hand, we are ready to consider the core postulates for revision:

R1 Closure: $T * \alpha = \mathrm{Cn}(T * \alpha)$
R2 Success: $\alpha \in T * \alpha$
R3 Inclusion: $T * \alpha \subseteq T + \alpha$
R4 Vacuity: If $\neg\alpha \notin T$, then $T * \alpha = T + \alpha$.
R5 Consistency: $T * \alpha$ is consistent if α is consistent.
R6 Extensionality: If $(\alpha \leftrightarrow \beta) \in \mathrm{Cn}(\emptyset)$, then $T * \alpha = T * \beta$.

As well as the supplementary postulates:

R7 Superexpansion: $T * (\alpha \wedge \beta) \subseteq (T * \alpha) + \beta$
R8 Subexpansion: If $\neg\beta \notin \mathrm{Cn}(T * \alpha)$, then $(T * \alpha) + \beta \subseteq T * (\alpha \wedge \beta)$.

Let's consider the basic idea of revision. We want to add a new belief to a belief set but cannot just expand by that belief since it results in a contradiction. Following this, Closure, Success, Inclusion and Vacuity are clearly desirable properties. Consistency is what distinguishes revision from expansion and Extensionality is common to all operators. The supplementary postulates are less obvious, even less than the corresponding contraction postulates. We should note that Superexpansion relies on the possible inconsistency of an expansion to cover the case where the reason that $T + (\alpha \wedge \beta)$ is inconsistent is that $T + \beta$ is inconsistent.

In the light of the Levi identity, we can construct maxichoice, full meet, and partial meet revision operators, with maxichoice always yielding a maximal theory and $T *_{full} \alpha = Cn(\alpha)$. Thus, just as with contraction, we generally only attend to partial meet revision operators.

The selection function of partial meet revision operators may be, via the Levi identity, defined in terms of various epistemic entrenchment relations with similar consequences: revision operators satisfy R1-R7 iff they are relational partial meet revision operators and satisfy R1-R8 iff they are transitively relational partial meet operators.

There have been many additions, alterations, alternatives, criticisms, etc. of standard AGM theory, but it remains the critical starting point for any discussion of theory change.

Chapter 3

Trivial Revision, Splitting Theories, and Parallel Interpolation

Classic AGM belief revision, and a good deal of subsequent work in the field, focuses on the revision of monolithic belief systems. The object of addition, contraction, revision, or update is the entire system (whether it is a belief set, belief base, or some other structure) and only a single system. In the standard idealization, the operators act without regard for relationships between belief *systems*. It is certainly worth investigating the relationship between distinct belief systems, for example, how addition in one belief system (my own) might relate to changes in another (a friend of mine). This may be as simple as updating my friend's beliefs when mine change. After all, my friend has beliefs about my beliefs. If my beliefs change, it is a change to the world and to the part of the world my friend has beliefs about. So this change to the world might fruitfully be modeled as an update. However, it is also plausible to see my friend and me as sharing beliefs, so that a change to this shared portion is just a revision (or simple addition) to our common store.

Similarly, we can consider belief systems that have significant substructure and want the revision operators to be sensitive to that structure. After all, we rarely make a change to our own beliefs as a whole, or wait for fully worked out equilibrium before considering additional changes. For example, when one edits a textbook, it is hard not to notice that some changes are *confined* to certain portions of a book because they are only relevant to the topic of some specific chapter. Other changes may require touching every part of the book. But some changes are inherently *local*.

Even the most hard-core justification holist or coherentist must admit to there being some sorts of structure in a belief set. For example, this is at the heart of the need to deal with trivializing contradictions in classical settings. The AGM revision postulates, as is well known, do not uniquely determine a revision operator. Instead, one must also choose an appropriate selection function over the remainders of a theory given a contradicting new belief in order fix the revision function. Standard AGM puts the priorities between elements of a belief set outside the belief set as if those considerations were extra-logical.

However, this independence of priority and logical structure is not characteristic of many forms of revision. Sometimes, logical structure matters. In particular, some parts of a belief set are more strongly interrelated than others. That is, belief sets may cover a range of *topics* and these topics may be strongly internally coherent with at best weak connections to various other topics. In disputation (or even ordinary conversation) it is commonplace to require disputants to "stay on topic". One risks fallacy if one tries to critique an argument by appealing to the lunacy of an opponent's beliefs in a entirely unrelated area. Forcing them to retract the unrelated belief, even for the sake of argument, should have little effect on the topic at hand. Of course, there is a rather large *ceteris paribus* clause: all other things being equal, lunacy in any area can cast doubt on otherwise unsupported beliefs. For example, testimony grounded primarily in the credibility of a witness can be weakened by evidence that the witness has verifiable false beliefs in other areas.

Staying on topic does not mean, in general, staying very narrowly on point, e.g., only appealing to facts that support a particular conclusion. What is germane to a discussion is subject to a wide variety of considerations, including methodological considerations, larger epistemic goals, and dialectical principles. An obvious example is if one is trying to learn what one's interlocutor knows about that topic.

In [Parikh, 1999], Rohit Parikh proposed an intuitively appealing approach to topic relevance and belief revision: revision should apply to the smallest relevant subset of the theory that can be "split" from the parent theory. Roughly, relevance is determined by whether the fragment and the input formulae have intersecting signatures, that is, whether they share basic vocabulary. What makes a fragment separable is that its signature is disjoint from the rest of the theory. For example, if we have a theory with sentences which mention great apes and other sentences which mention paper pulp, but there are no sentences which mention both great apes and paper pulp, then we could hope that these topics and their associated sentences could be separated into disjoint subtheories. [1]

Parikh justifies his approach in three ways:

1. He offers a rationality critique of the standard AGM revision postulates, to wit, that they sanction trivializing revision operators. Such operators necessarily destroy information that is (topic) unrelated to the new belief.

2. Since the portion of a belief set that is (topic) relevant to the new belief can be smaller than the entire belief set, it is possible that topic-oriented revision is (com-

[1] Obviously, this disjointness needs to be made precise since in first order logic, it is easy to find sentences that trivially bring signatures together. Thus, if a belief set contains the sentence "Turtles are not great apes", then it will also contain the sentence "Turtles are not great apes or paper pulp is used to make paper." See section 3.2 for a lengthy discussion.

putationally) *easier* than standard revision.

3. Finally, Parikh claims that split-based revision corresponds to (most) actual revision, and thus is more realistic.

The first and third considerations are related and in tension. AGM theory is an idealization and, as such, when testing it against practice or against intuition, we must be careful not to import features that are irrelevant to the idealized context. That is, if we are trying to determine what is "fully" rational by asking what an ideal agent *free* of resource constraints would do, it doesn't make sense, in the same breath, to require said agent to adhere to rubrics we impose on resource-bound agents. Requiring such rubrics inverts the idealization method. We idealize in order to help justify the rationality of resource-bound principles as either (the best, or the most feasible) *approximations* of the ideal ones, or, thinking consequentially, as most likely to reliably lead to (at least, reasonable approximations of) the ideal *outcomes*. This picture can be made considerably more complicated. For example, instead of measuring the rationality of principles designed to guide resource-constrained (or actual!) agents solely by how closely they approximate the principles guiding an ideal agent, we could claim that at least some of the resource-sensitive principles are independently rational. This need not entail rejection of ideal agent methodology, though, obviously, the more situated the rationality considerations, the more indirect the connection to ideal agents. Furthermore, there are many ways of getting away from approximation; for example, one could justify the rationality of resource-sensitive principles by showing that they are principles that an ideal agent would *advise* us to adopt in the given situation.

Since we are, ourselves, not ideal agents, we have to make some judgments about what should be preserved in the ideal context. If the only reason that we engage in, or find natural, topic sensitive revision is that it is computationally easier, then there is not a strong case that it is a core aspect of rational revision, simply because we could discard it if we had unlimited resources. On the other hand, we might regard some computational boundaries as more inherent. To pick the most obvious example, we might treat decidable problems, semi-decidable problems, and undecidable problems as corresponding to different sorts of idealized agents: e.g., agents working with arbitrary finite resources or agents working with arbitrary and infinite resources. Most of the time when reasoning about algorithms and programs we do confine ourselves to computable functions (i.e., to universal Turing machines) as the natural ideal agent even though we know that all our realizations of programs will be quite sharply bounded in time and space. (Clearly, there are many situations where we want to have those bounds firmly in mind, especially when faced with large effects due to, e.g., the *kind* of space — registers, cache, main memory, or disk — we're dealing with. When trying to empirically evaluate an algorithm with certain theoretical complexity properties, we have to take care that our abstraction from the physical situation doesn't obscure confounding factors.) It is plausible to appeal to actual practice and intuitiveness when formulating rationality criteria as a buttressing argument. All things being equal, a model of rationality that comports well with what real agents can do and recommend is superior since it requires less departure from actual norms and thus less of an explanatory gap between what we believe ought to, ideally, guide us and what *works*.

One way to distinguish the rationality argument from the naturalness argument is in what they recommend. A strong enough rationality point can directly require a change to the revision postulates themselves (which is what Parikh argues). In other words, rationality determines a set of permissible revision operators. Assuming we accept the postulates, naturalness can help us select *which* rational operator to use. If the set of rational operators contains some natural (or actual) ones, this buttresses the correctness of the postulates. Conversely, if no natural operator is sanctioned by the postulates, we have to either adjust the postulates or have an explanation for why the natural operators are irrational. Such explanations range from an approximation account (wherein natural revision operators compromise on rationality to some degree for a variety of historical, evolutionary, or resource determined reasons) to a full blown error theory a la Mackie [Mackie, 1977]. After all, people may simply be systematically irrational. People are certainly quite bad at all sorts of reasoning.

However, we must take care not to go too psychologistic. People might be quite bad at adhering to various rational norms, but rationality is, in the name, a norm for and of people. Naturalness of a revision operator is not merely a matter of whether the operator is psychologically realistic, something which most people would not be able to assess. A natural revision operator is one that is recognizably related to our general practice of assessing rationality both theoretically and in practice. In particular, we must be persuadable that the sanctioned operators are reasonable for us to adopt (at least as ideals).

As we will see, the direct rationality case Parikh makes (which dominates his argument) is weak, so we will need to revisit and expand the other arguments.

3.1 The Problem of Trivial Revision

Parikh's critique of the standard AGM postulates is that that they are consistent with *trivial revision*. Parikh defines the trivial revision[2] operator as follows:

> If [a formula] A is consistent with [a belief set] T, then $T * A = T + A$, otherwise $T * A = [\mathrm{Cn}](A)$. [Parikh, 1999]

In other words, a trivial revision, in case of conflict, is simply the replacement of the revised belief set with the new information. It is certainly the case that such a revision operator is consistent with the core AGM postulates for revision. Furthermore, the risk of such a revision operator well known in the literature. For example, consider the discussion of full meet contraction in [Alchourrón et al., 1985]:

> [I]t is tempting to try the [contraction] operation $A \sim x$ defined as $\bigcap(A\perp x)$ when $A\perp x$ is nonempty, and as A itself in the limiting case that $A\perp x$ is empty. But as is shown in Observation 2.1 of [2], this set is in general far too *small*. In particular, when A is a theory which $x \in A$, then $A \sim x = A \cap \mathrm{Cn}(\neg x)$. And thus, as noted in Observation 2.2 of [2], if revision is introduced as usual via the Levi identity as $\mathrm{Cn}((A \sim \neg x) \cup \{x\})$, it reduces to $\mathrm{Cn}((A \cap \mathrm{Cn}(x)) \cup \{x\}) = \mathrm{Cn}(x)$ for any theory A and proposition x inconsistent with A. In other words, if we revise a theory A in *this* way to bring in a proposition x inconsistent with A, we get no more than the set of consequences of x considered alone

[2]Parikh actually calls this trivial *update*, but I prefer to maintain the distinction between revision and update a la [Katsuno and Mendelzon, 1991].

— a set which is far too small in general to represent the result of an intuitive process of revision of A so as to bring in x.

The *locus classicus* of the standard AGM postulates acknowledges that the trivial revision operator is permitted by the core revision postulates, but instead of modifying the postulates, it mobilizes the intuitive irrationality of trivial revision as a justification of partial meet contraction (and yet, partial meet contraction contains trivial revision as a limit case). The standard development is repeated several times (e.g., see [Gärdenfors, 1988]):

1. Argue for the core contraction postulates.
2. Note that contraction postulates do not uniquely determine an operator (unlike expansion).
3. Explicate the remainder operation
4. Consider in turn maxichoice (too big), full meet (too small), and partial meet contraction (what's left, and a generalization of the other two).[3]
5. Show how to generate sensible partial meet contraction operators using epistemic entrenchment.

In this dialectic, that the core postulates alone do not rule out trivial revision (or revision that always results in a complete theory independently of whether the original belief set was complete) is not a problem with the postulates. Obviously, in general, other considerations are required to narrow down contraction. A *choice* must be made. Partial

[3]This Goldilocks story is repeated several times in the literature.

meet contraction locates this choice in the selection function — if your selection function returns, at best, a singleton, then you have maxichoice contraction; if your selection function returns all the remainders, then you have full meet contraction. These form the upper and lower bounds of rational contraction operators, though they are not, typically, themselves sensible. Of course, there are specific cases where a given, sensible selection function intersects with a maxichoice or a full meet selection function, at least, for some particular pair of theory and proposed revision. One must be careful to distinguish between the odd trivial (or maximizing) revision, and a revision operator that does *nothing* but trivial (or maximizing) revisions. But there are lots of selection functions of dubious rationality. For example, instead of returning the maximally entrenched (according to some entrenchment relation) elements of $A \perp x$, a selection might return a *random* subset. Obviously, such a contraction operator can be critiqued for being somewhat arbitrary and for ignoring critical information about $A \perp x$. It seems a bit much to expect the postulates alone to rule out such an operator, along with an arbitrary number of similar ones. At least, it's hard to see why new *postulates* are required, instead of a simple discussion of selection functions.

Perhaps more importantly, while the general irrationality of this example is plausible, it also seems to be the case that it is sometimes sensible. For example, if one is concerned that a given entrenchment relation is biased or otherwise problematic, a random selection function might be more reliable overall. Similarly, there could be contexts where a trivial or a maximizing revision operator was sensible. In general, one could regard the postulates as giving necessary conditions for rational operators, not sufficient ones. Indeed, as AGM theory aims to be a fairly general account of revision (useful for modeling anything from people changing their mind, to communities evaluating legal changes) we should *expect* and *prefer* it to err on the side of liberality.

Before delving into the rationality of trivial revision, we shall examine Parikh's approach to restricting it. It is curious that maximizing revision receives no consideration from Parikh, especially since the lack of rationality of maxichoice is given equal attention in the literature.

Perhaps maxichoice is not as intuitively or severely irrational to him. It certainly is *odd* and odd in a symmetric way to full meet. Maxichoice revision has the bizarre consequence that every revision is a complete theory (that is, for every formula ϕ, it includes either ϕ or $\neg\phi$. For Parikh the core motivation is avoiding the loss of *unrelated* information:

> The existing set of beliefs T may contain information about various matters. E.g. my current state of beliefs contains beliefs about the location of my children, the state of health of my teeth, and beliefs about the forthcoming election in India. In case one of my beliefs about the location of my children turns out to be false, it surely ought not to affect my beliefs about the election, since the subject matters of the two beliefs do not interact in any way.

Thus, for Parikh, the problem of trivial revision is specifically a problem of *topic* relevance. That is, it is the lack of interaction — that is, the disjointness — of the *subject matters* of the beliefs that is key. However, this does not explain the neglect of maxichoice revision. Presumably, *adding* information about irrelevant topics is as bad as removing some.

3.2 Splitting Theories

One obvious tactic for modeling this disjointness would be to partition the belief set into disjoint subsets. However, there are a number of problems with this move. Consider the case of the belief set, $T = \mathrm{Cn}(\{P, Q\})$. Clearly, there is at least one partitioning of T into two parts, T_1 and T_2 such that $P \in T_1$ and $Q \in T_2$. But since T_1 and T_2 *partition* T, they must be disjoint. But then it is unclear where to put $P \vee Q$ or $P \wedge Q$, both of which are elements of T. Clearly, there are (at least) four distinct (families of) partitions with regard to these two compound formulae (depending on where we put the disjunction and conjunction). The elements of the partitions cannot be both disjoint *and* deductively closed themselves (since adjunction will have to be blocked in one of them), and there is an unsettling arbitrariness to the division. In particular, it is hard to say that the partition divides the formulae of T into sets of formulae which *do not interact* with formulae in another partition. Obviously, there are inferential relations that can cross partitions. Indeed, for *any* partitioning of a deductively closed set into more than one partition, there will be formulae in some partition that are entailed by formulae in a distinct partition (and thus in the deductive closure of that partition).

First, if we consider tautologies, then it is trivial, since any tautology is entailed by an empty set of formulae (and given that our formalism is monotonic), then no matter which element of the partition a particular tautology goes in, it will be entailed by each formula in all the other elements. Second, if the belief set is inconsistent, then since it contains all formulae, it will contain the self contraction $P \wedge \neg P$, and no matter which element we put that in, it will entail all the other formulae, thus some formula in another element. But perhaps these are cheats, since tautologies do not really have a subject matter at all and standard AGM revision presumes consistent belief sets (by and large).

Third, suppose you have a partition of the contingent formulae of a consistent T, $\{T_1 \ldots T_n\}$. Let T_i be an element of the partition such that for all $f_i \in T_i$, there is no $f_j \in T_j$ (for any $j \neq i$) such that $f_j \models f_i$. $(f_i \vee f_j) \in T$ since $f_i \in T$ and T is deductively closed. Suppose $(f_i \vee f_j) \in T_i$. Then, since $f_j \models (f_i \vee f_j)$, there is at least one element of the partition (namely, T_i) which contains a formula entailed by a formula in a distinct element. If $(f_i \vee f_j) \notin T_i$, then it is in some other element of the partition (since the partition is exhaustive), say T_j. But then T_j contains a formula entailed by a formula in a distinct element.

Just partitioning the formulae, even neglecting the tautologies, fails to yield a non-inferentially-interacting partition. If we consider again the initial example where $T = \mathrm{Cn}(\{P, Q\})$, it is clear that some formulae interact directly (e.g., $P \models (P \vee Q)$, $Q \models (P \vee Q)$) and some interact only indirectly (e.g., $P \not\models Q$ and $Q \not\models P$, but $\{P, Q\} \models (P \wedge Q)$). For example, formulae which are either an atomic letter or the negation of an atomic letter (i.e., the literals) do not directly interact with literals with a different atomic letter (that is, with literals with a different signature) but require some other formula to mediate their interaction. This follows fairly quickly from Craig's Interpolation. In any consistent T, for any pair of complementary literals (e.g., $\{A, \neg A\}$) only one can be an element of T. Literals fix the basic facts about the world, and sets of literals could be thought of as a subject matter. Some belief sets follow from a set of (consistent) literals alone, obviously, such as $\mathrm{Cn}(\{P, Q\})$ or $\mathrm{Cn}(\{P, (Q \wedge R)\})$, but some do not, such as $\mathrm{Cn}(\{P, (Q \vee R)\})$. In the latter case, the set $\mathrm{Cn}(\{P, (Q \vee R)\}$ entails neither Q nor R, and is not entailed by $\{P, Q, R\}$.

These points lead naturally to Parikh's splitting methodology. Instead of literals, per se, he first partitions the *signature*[4] of the logic (or simply of the belief set in question). In the propositional case, a partition will be a set of pairwise disjoint sets of propositional variables. A partition *splits* a belief set (aka theory) T if we can find formulae in T whose signatures each are subsets of a distinct element of the partition and those formulae, together, entail T. More formally:

Definition 4. (From [Parikh, 1999]:) *For a theory T with a signature S, a partition of S,* $\{S_1, \ldots, S_n\}$ splits *T if $T = \mathrm{Cn}(A_1, \ldots, A_n)$ and for all i, formula A_i is in S_i. The partition is a splitting, or, more specifically, a T-splitting.*

It is possible for $\mathrm{Cn}(A_1) \ldots \mathrm{Cn}(A_n)$, even excluding tautologies, to have non-empty intersections. Indeed, if the consequences are with regard to the entire signature of T, then they must have non-empty intersections (since for each element in some $\mathrm{Cn}(A_i)$, one can always form a disjunction with some element of $\mathrm{Cn}(A_j)$ where $j \neq i$). But, if $\mathrm{Cn}(A_i)$ defined over the signature of A_i, which is disjoint with that of every other A_j, then $\mathrm{Cn}(A_i) \ldots \mathrm{Cn}(A_n)$, excluding tautologies[5], will all be likewise disjoint. Thus, $\bigcup(\{\mathrm{Cn}(A_i), \ldots, \mathrm{Cn}(A_n)\}) \subset \mathrm{Cn}(A_1, \ldots, A_n) = T$.

It seems that Parikh's notion of a splitting does fit in with the idea that there may be distinct subject matters in a theory which are in some way independent. For a belief

[4]Parikh calls them "languages" in [Parikh, 1999], but that is a bit confusing since "languages" seem more naturally, in this context, to refer to the set of well formed formulae of a grammar with regard to a specific signature.

[5]Though, again, if the tautologies are expressed only in the signature of A_i, they will be disjoint from the tautologies of the other partitions.

set, we can capture all the significant beliefs with regard to a subject matter (that is, a signature) if we find beliefs which *generate* all the other beliefs about *nothing but* that subject matter. The generating set of beliefs then, themselves, can be divided according to their (joint) signatures. There will be quite a few *cross cutting* beliefs which are part of the theory but not part of any of these generated sets. These "extra" beliefs are not *free standing* in the sense that they are required in order to capture aspects of the totality of the theory. That is, we can create a generating base for the entire theory without having to incorporate these formulae or *any* formula which has an identical signature. This is exactly the sense in which they are extra...they are not required in order to completely characterize the theory's commitments to a (disjoint) topic.

Alone, this might not make for a compelling purely logical model of subject matter or "topic". In general, there are many ways to split a belief set. Consider the following example:

Example 4. *Let $A =$'Roses are red'; $B =$'Violets are blue', $C =$'Sugar is sweet'.*

Let $T = \mathrm{Cn}(A \wedge B \wedge C)$.

The sets $\{\{A, B\}, \{C\}\}$, $\{\{A, C\}, \{B\}\}$, $\{\{B, C\}, \{A\}\}$, and $\{\{A\}, \{B\}, \{C\}\}$ all split T (e.g., as $\mathrm{Cn}(A \wedge B, C)$, $\mathrm{Cn}(A \wedge C, B)$, etc. respectively).

Intuitively, $\{\{A, B\}, \{C\}\}$ seems more naturally topic oriented than the others, but only because of external information that roses and violets are flowers and red and blue are colors. These connections are, of course, opaque to the representation, thus provide no purely logical grounds for preferring the $\{\{A, B\}, \{C\}\}$ split over the others.

Fortunately, there is a purely structural criterion that distinguishes splittings: the granularity of the splitting. In example 4, some splittings are not maximal, that is, there is another splitting which, in Parikh's terminology, *refines* those splittings. $\{\{A\}, \{B\}, \{C\}\}$ refines $\{\{A, B\}, \{C\}\}$ because some element of the latter (i.e., $\{A, B\}$) can be further subdivided into (disjoint) elements of the former (i.e., $\{A\}$ and $\{B\}$), and both are splittings of T. Parikh establishes (for the propositional, finite signature case) that there is always a unique finest splitting. The elements of such a splitting cannot be further subdivided while remaining a splitting.

Example 5. *Let A =‘Roses are red’; B =‘Violets are blue’, C =‘Sugar is sweet’.*

Let $T = \mathrm{Cn}((A \vee B) \wedge C)$*. The set* $\{\{A, B\}, \{C\}\}$ *splits T and it is the unique maximal splitting. It is also the sole splitting (other than the identity splitting).*

Let us pause for a moment to recognize that the fact of there always being a unique finest splitting is rather a striking one. Even recognizing that a finest splitting might be, in some sense, too fine (as in example 4, where the content of A and B, against our general background knowledge suggests co-topicality), the univocality is compelling. The elements of a finest splitting *cannot* be further split. They are logically cohesive. The flip side of their cohesion is their separability from other elements of the splitting. While there will be formulae in a belief set T that are only entailed (and thus only in T) by the combination of several of the splitting formulae, the key point is that to prove them, you do not need formulae whose signature intersects the signature of more than one element in the *finest* splitting. That is, there are no required formulae which essentially span elements of the finest splitting. The "extra" formulae of the finest splitting are, in principle, inherently extra. The splitting isolates the way basic vocabulary is used in the belief set. (In section 3.3, this point is highlighted in Kourousias and Makinson's notion of *parallel interpolation*.)

These two components are critical to the splitting account of topicality. Any splitting separates parts of the signature and it certainly seems that separability is necessary for non-topicality. If you can separate two terms of the signature and preserve the theory, then the theory does not rely on any connection between those terms. Finest of granularity takes that further: if the theory doesn't rely on a connection between two terms, then they should be separated. If there were multiple distinct finest splitting, then (logical) topicality would be indeterminate: it would depend on some extra logical choices one made. Univocality of the finest splitting tells us that there is a purely logical topic structure in all theories.

To tie this back to revision, Parikh proposes to use splitting to rule out trivial revision. Intuitively, what's wrong with trivial revision is that it touches irrelevant parts of the belief set. "Irrelevance" is characterized by disjointness of signature. If the signature of an element of a splitting does not intersect with the incoming formula, the revision operator should not affect formulae entailed by the characteristic formula of that element alone. If $T = \mathrm{Cn}(A, B)$ where A and B are respectively expressed in the signature of a finest splitting of T, $\{T_A, T_B\}$, then, where $\mathrm{Sig}(\alpha) \cap T_A = \emptyset$, $\mathrm{Cn}(A) \subseteq T * \alpha$. Parikh expresses this as an additional revision axiom to the core AGM axioms:

Axiom P. *(From [Parikh, 1999]:) Let* T *be split by* $S_T = \{S_1, \ldots, S_n\}$ *and let* C *be an arbitrary formula. Let* $S_C = \bigcup(\{S \in S_T \mid \mathrm{Sig}(C) \cap S \neq \emptyset\})$*. Let* $SF_T = \{SF_1, \ldots, SF_n\}$ *be formulae such that* $T = \mathrm{Cn}(SF_1, \ldots, SF_n)$ *and* $\mathrm{Sig}(SF_i) \subseteq S_i$ *for all* $1 \leq i \leq n$*. Let*

$SF_C = \{SF \in SF_T \mid \mathrm{Sig}(SF) \cap S_C \neq \emptyset\}$.

Then $T * C = (\mathrm{Cn}(SF_C) *_i C) + (\{SF_1, \ldots, SF_n\} \setminus SF_C)$*, where* $*_i$ *is the update operator for* L_i*.*

Essentially, Axiom P states that actual revision must be performed only on the components of the split theory that are related, by their signature, to the input formula.

It is easy to see that Axiom P blocks triviality in many cases. Even if $*_i$ itself is trivializing, it will only affect A_i, leaving the rest of the theory intact. In fact, it is fairly plausible to make $*_i$ trivializing in many circumstances, especially if the splitting is very fine: Upon receiving new information about a topic that contradicts current beliefs on that topic, one might find oneself rejecting all of one's prior beliefs on that topic on the grounds that if one is wrong on that point, perhaps one is not trustworthy on that topic at all. Similarly, if there are several ways to resolve the contradiction, one might not feel in a position to determine which is the correct resolution. It may be reasonable to suspend judgment on those matters until one can perform a more systematic review of that subject area. (We shall take up this point below.) Of course, if there is no splitting of T finer than T itself, then $*_i = *$ and thus $*$ can be trivial in some cases.

So, is Axiom P a solution to "the *problem* of trivial revision"?

Axiom P certainly improves the preservation properties of the core AGM axioms, by blocking many trivial (or wanton) revisions. It does not prevent revision operators from either sometimes being trivial, or from being trivial *within* the scope of a topic. The former is perhaps less worrisome. It hardly seems irrational that *sometimes* a revision operator will throw out everything — sometimes new information does in fact conflict (strongly) with our prior beliefs. Perhaps the revised belief is central, or our belief set is relatively impoverished. In these cases, we might expect the revision to be extreme. In fact, splitting gives us a way of determining when a new belief is strongly conflicting with the whole of our current beliefs, thus yielding operators that are (globally) trivializing only when justifiably so.

However, localized trivial revision seems intuitively problematic. A local revision operator $*_i$ can be relentlessly trivializing for its topic. Indeed, we could define $*_i$ by full meet contraction. Nothing in Axiom P rules that out. But, to repeat from [Alchourrón et al., 1985] :

> In other words, if we revise a theory A in *this* way [i.e., via full meet contraction/trivial revision] to bring in a proposition x inconsistent with A, we get no more than the set of consequences of x considered alone — a set which is far too small in general to represent the result of an intuitive process of revision of A so as to bring in x.

This seems no less true when revising a topic coherent subpart of A with a topic specific belief x. We know that there are operators that will preserve more! It seems that localizing revision only solves the problem if the only reason that trivializing revision was irrational was because it eliminated irrelevant beliefs. Axiom P mitigates trivial revision, but it does not eliminate the problem. If one has to appeal to considerations beyond the axioms *anyway* to rule out locally trivializing operators, why not do so in the global case?

Consider the following case:

Example 6. *Let* $T = \mathrm{Cn}(A, B \to \neg A, C)$*.* T *can be split by* $\{\{A, B\}, \{C\}\}$*.* $T * B$*, if* $*$ *is* *locally trivializing, will be* $\mathrm{Cn}(B, C)$*. However, now neither* $B \to \neg A$ *nor* $\neg A$ *nor* A *are in* $T * B$*. We have lost all information about* A*.*

While in this example, A and B are connected and thus, logically speaking, are related, it seems completely wanton to obliterate all our prior understanding of A when adding B to our knowledge. While not *as bad* as a globally trivializing revision (since at least C is safe), it certainly seems bad enough to demand some justification.

Parikh might respond that Axiom P has some inherent advantages that are sufficient to justify it even if it does not entirely block trivial revision. Adopting a splitting approach is, Parikh argues, computationally helpful as the revision operation is performed on a subset of T, perhaps a small one (though one must deal with the consequences of computing the splitting). The idea of revising only the "relevant" portion of a belief set is very natural in a wide variety of settings. Axiom P conforming operators *are* better in that they do protect logically irrelevant information. So entrenchment relations can be more topic specific and thus, one can hope, a bit nicer, or even more stable between revisions.

However, Axiom P does not come for free. Parikh argues that the supplementary AGM axioms, 7 and 8 should be discarded:

> We do not feel that these axioms [7 and 8] are consistent with the spirit of our work for the following reason. Suppose that $A = (\neg P \vee Q)$ and $B = (P \vee Q)$, then $A \wedge B$ is equivalent to Q and says nothing about P. Now revising a theory T first by A could cause us to drop some P-related beliefs we had, and revising after that with B we might not recover them. But revising with $A \wedge B$ should leave our P beliefs unchanged, provided that our beliefs about P

> and Q were not connected. Thus contrary to 7, revising with the conjunction may at times preserve more beliefs than revising first with A and then with B. This is why it does not seem to us that axioms 7 and 8 *should* hold in general.

Giving up 7 and 8 breaks the connection between the revision axioms and transitively relational partial meet contraction functions (given by the representation theorem 4.16 of [Gärdenfors, 1988], pg. 82). In general, axioms 7 and 8 (and their corresponding contraction postulates) correspond to ordering conditions on the remainders, that is, to conditions on the epistemic entrenchment relations. While axioms 7 and 8 may be a bit counterintuitive on their face, the transitively relational orderings of remainders seems *very* natural.

This de-emphasis of the operators is characteristic, as George Kourousias and David Makinson point out in [Kourousias and Makinson, 2006]:

> Parikh and collaborators have studied the problem of modifying the notion of partial meet revision so as to ensure that it satisfies the relevance criterion. They have done so *syntactically*, i.e., by exampling what further postulates may be added to the standard ones of AGM to ensure respect of relevance…However one may also approach the situation from a more *semantic* angle, asking how we might modify the definitions of partial meet contraction and revision so as to ensure that they respect relevance.

Since they approach localized revision from the semantic point of view[6], their ac-

[6]Though, we quibble a bit with their notion of "semantic". The contraction and revision operators seem just as syntactic as the postulates. Model theoretic accounts seem much more semantic.

count might illuminate the problems discussed above.

3.3 Normalized Revision

Parikh starts from the core AGM axioms and asks what can be added to rule out (certain forms of) trivial revision. Kourousias and Makinson start from partial meet contraction and try to find under what conditions it avoids (certain forms of) trivial revision. They characterize the problem of trivial revision in terms of it removing "irrelevant" beliefs, where relevance is defined in terms of finest splittings:

Definition. *(Parikh-)Irrelevant to a belief change (Viz def. 3.1 in [Kourousias and Makinson, 2006].) A formula, $\beta \in T$,* is (Parikh-)irrelevant to *$T * \alpha$ (or $T - \alpha$) iff for the unique finest splitting of T, $S_T = \{S_1 \ldots S_n\}$, the intersection of* $\mathrm{Sig}(\beta)$ *and* $\bigcup(\{S \in S_T \mid S \cap \mathrm{Sig}(\alpha) \neq \emptyset\})$ *is empty.*

Irrelevant formulae, on this definition, are not just formulae that have nothing in common with the input, but also have nothing in common with any formulae that are relevant to the input. Note that relevant formulae need not have an overlapping signature with the input formula — just a signature that is connected to that of the input (e.g., via a chain of overlappings).

Example 7. *Let* $T = \mathrm{Cn}(A \rightarrow B, B \vee C, C \vee D, E \vee F, F \vee G)$. *The finest splitting is* $S = \{\{A, B, C, D\}, \{E, F, G\}\}$. *Consider the contraction* $T - (A \rightarrow B)$. *In this case,* $C \vee D$ *is relevant even though the respective signatures of the formulae are disjoint, since their signatures both intersect with a common element of the finest splitting. Similarly,*

$(E \vee F) \wedge (B \vee C)$ *is relevant to this change.* $E \vee F$ *is not relevant to* $T - (A \rightarrow B)$*, but is to* $T * G$ *(though, in this case, we just have an expansion by* G*).*

Clearly revision operators which conform to the AGM core axioms + Axiom P respect this sort of relevance, since Axiom P explicitly constructs revision operators to add in all the "irrelevant" formulae. Kourousias and Makinson show that you can respect Parikh-relevance by massaging the belief set into a specific form, then applying standard partial meet contraction (and revision) upon that. The normal form in question is a belief base which consists of a set of generating formulae which characterize the finest splitting of a theory. That is, a set of sentences each of which have maximally disjoint signatures and which collectively entail the original theory.

Definition 5. *A* split base, $\{SF_1, \ldots, SF_n\}$, of *a set of sentences,* SF*, is a (maximally) split base of a (consistent) belief set* T *iff for a (finest) splitting of* T*,* $S_T = \{S_1, \ldots, S_n\}$*,* $\mathrm{Sig}(SF_i) \subseteq S_i$ *for all* i *such that* $1 \leq i \leq n$ *and* $T = \mathrm{Cn}(SF_1, \ldots, SF_n)$*.*

For any splitting of T there is an equivalence class of (maximally) split bases of T. Kourousias and Makinson establish the following theorem:

Theorem. *Split base contraction respects Parikh-relevance*

(From [Kourousias and Makinson, 2006], theorem 4.1.)

Let T *be a belief set and let* SF_T *be a maximally split base of* T*. For any partial meet contraction operator,* $-$*, if* $\beta \in T$ *is [Parikh-]irrelevant to* $T - \alpha$*, then* $\beta \in SF_T - \alpha$*.*

(The corresponding theorem for revision follows immediately from the Levi identity.)

In their informal discussion of this theorem, Kourousias and Makinson point out that the reason that partial meet contraction against the split base respects Parikh relevance is that when constructing the remainders of T with regard to α (i.e., $T \perp \alpha$), all formulae in "irrelevant" elements of the splitting have to go into each remainder. Essentially, the normalization process removes from consideration all sentences that are inessential, thus, all "trivial" material conditionals. Such conditionals force there to be signature overlap chains between otherwise unrelated sentences.

Example 8. *In this example, we examine three representations of the same belief set: the belief set itself, a generating belief base, and a split base with respect to the belief set:*

- $T_{base} = \{A, A \to B, C \to B\}$
- $T = \mathrm{Cn}(T_{base})$
- $T_{nbase} = \{A, B\}$ *(the split base of T).*

  Note that it makes a difference whether we consider the language of T_{nbase}, L_n, to include C or not. If $C \notin L_n$, then $T \neq \mathrm{Cn}(T_{nbase})$. The splitting language, in this case, should be $\{\{A\}, \{B\}, \{C\}\}$. We can force C into either other element, but then the choice is arbitrary.

Each of these representations has a distinct set of remainders with respect to B:

- $T \perp B = \{\mathrm{Cn}(A \to B, C \to B),$

  $\mathrm{Cn}(A \to B, \neg C \to B),$

  $\mathrm{Cn}(A, \neg A \to B, C \to B),$

  $\mathrm{Cn}(A, \neg A \to B, \neg C \to B)\}$

- $T_{base} \perp B = \{\{A, C \rightarrow B\}, \{A \rightarrow B, C \rightarrow B\}\}$

- $T_{nbase} \perp B = \{\{A\}\}$

Clearly, (non-empty) selection functions on these remainders will vary in their results by quite a bit. For example, $T_{nbase} \perp B$ has only one possible selection (and thus, the intersection of the elements of the selection will just be $\{A\}$ again). No selection from either $T \perp B$ or $T_{base} \perp B$ can be $\{A\}$, or even $\mathrm{Cn}(A)$. Furthermore, there is no intersection of any of those selections which is equal to $\{A\}$. Thus, there is no (non-trivial) revision common to all these remainders. Furthermore, there are selection operators on $T \perp B$ and $T_{base} \perp B$ which generate revision operators that do not respect Parikh-relevance (i.e., any revision which does not contain A, which are easy to generate from $T \perp B$ and T_{base}).

Notice that both in T and in T_{base}, A is entangled with B in an inessential way via $A \rightarrow B$ (i.e., $\neg A \vee B$). T_{base} illustrates some of the consequences of not taking care to avoid $A \rightarrow B$. The (somewhat redundant) base $\{A, B, C \rightarrow B\}$ has only a single remainder, $\{A, C \rightarrow B\}$, and as a result, all of its (non-trivial) revisions contain A. Interestingly, those revisions are not the same as those of the split base.

In example 8, the non-split revision operators include some which preserve formulae (when revising by $\neg B$) that are lost by the split ones, especially those formulae with C in their signature, even though those formulae do not entail B. It seems quite reasonable to allow for the possibility that $(C \rightarrow B) \in (T * \neg B)$ or $(C \rightarrow B) \in (T_{base} * \neg B)$. The situation is perhaps starker if we consider a case with no freestanding propositional variables:

Example 9. *Again, consider three representations of the same belief set: the set itself, a generating base, and a split base:*

- $T_{base} = \{A \vee B, C \vee B\}$
- $T = \text{Cn}(T_{base})$
- $T_{nbase} = \{(A \vee B) \wedge (C \vee B)\}$

Now let us compare the remainders with respect to $A \vee B$ (for simplicity, we just display the remainders for the two bases):

- $T_{base} \perp (A \vee B) = \{\{C \vee B\}\}$
- $T_{nbase} \perp (A \vee B) = \emptyset$

$T \perp (A \vee B)$ will have elements containing formulae such as $C \rightarrow (A \vee B)$, that is, various conditionals generated off the elements in the generating base, with both positive and negative antecedents.

Since B is connected to both A and C, there is no finer splitting than the entire signature of T. As we see in this case, we can lose more information by performing normalized revision than straight (even base) revision. This seems to be a devastating blow against Parikh's professed motivation for Axiom P and the whole splitting move. While $(C \vee B)$ is related by topic to our input formulae, it is not *inferentially* related to $(A \vee B)$. Whether we use the AGM axioms alone, or we augment them with Axiom P, the set of conforming operators will contain overly aggressive operators.

We should also notice that this is a case where the Axiom P formulation and the split base formulations yield *different* sets of operators. In example 9, Axiom P sanctioned operators are just the normal partial meet operators (including full meet!) since the finest splitting is the original base. However, normalizing allows for only trivializing operators.

Thus, Axiom P (and similar) does not rule out trivial revision in all cases, and some of the cases where it fails to rule out trivial revision are not degenerate, that is, trivial revision would result in loss of information which is plausibly irrelevant to the change. Axiom P also blocks the supplementary postulates which, given the representation theorems, means that it blocks transitively relational operators. Whatever your attitude to the supplementary postulates, they are not completely bonkers. Gardenförs argues extensively for them, e.g., in [Gärdenfors, 1988], especially pp. 55-58. There, he motivates them as a way of capturing an intuitive constraint on iterative revision, gives a possible worlds justification, and explores the "useful consequences" of adopting them. Immediately thereafter, he discusses "three principles that have a certain prima facie plausibility but that must be rejected as general postulates for revision functions" ([Gärdenfors, 1988], pg. 58). The supplementary postulates are independently motivated, but then gain a great deal of plausibility from their correspondence to conditions on the entrenchment relation. He later develops a *sui generis* account of epistemic entrenchment ([Gärdenfors, 1988], pp. 86-91) which is closely related to the full set of AGM postulates. Scorning the supplementary postulates must be done with care.

Obviously, if Axiom P does not block trivial revision in the case where the finest splitting corresponds with the whole belief set, it likewise does not block trivial revision of a particular partition even when the number of partitions is greater than one. This is so obvious that Parikh uses it as an example of a revision operator that meets his modified postulates in a proof!

> [I]f A is not consistent with T, then write $T = Con(B, C)$ where B, C

are in L^T_A[7], $L - L^T_A$. Then let $T * A = Con(A, C)$...In this update procedure we used the trivial update on the sub-language L^T_A, *but we did not need to*. [emphasis added]

We need not think that this is a *paradigm* operator; it was just convenient for Parikh's existence proof (i.e., that the set of operators conforming to the core postulates plus Axiom P was not empty). However, given his extreme distaste for global trivializing revision operators, this blasé acceptance of a locally trivializing operator is a bit off. Surely it should not go *without comment* that this local trivialization is fine whereas global trivial revision is anathema (though not really, since Axiom P rules it out anyway). If core AGM + Axiom P *only* sanctioned local trivial revision, would this be an acceptable outcome? Would this be a more acceptable outcome than normal AGM sanctioned operators?

Clearly the answer to both questions is "no", at least from the perspective of minimizing information loss. The only reason given (thus far) to add Axiom P is to rule out operators which remove information that could have been preserved (i.e., in disjoint splits). However, if AGM + Axiom P only sanctioned local trivial revision, then there would be operators that it ruled out which preserved more information than the sanctioned ones. While we can argue (as we do below) that locally trivial revisions, or even globally trivial revisions, are sometimes justifiable, it is hard to see that they are *required*. At least, such a requirement is not compatible with any recognizable notion of informational economy, as it forces one to always select revisions that throw out more, sometimes much more, than is strictly necessary. As we have already noted, it is standard to mark

[7] I.e., the partition of L such that $\mathrm{Sig}(A)$ is a subset of it.

trivial revision operators as unacceptable operators except as a lower bound. It is hard to see how Parikh can appeal to this sort of move *only* in the topic sensitive case. If it justifies locally trivial revision operators, it justifies global ones as well. More precisely, the move is to argue that the inclusion of trivial revision operators in the set of AGM sanctioned operators does not show that there is a problem with AGM theory, merely that the postulates do not capture every aspect of the rationality of revision.

This does not mean that topic oriented contraction and revision based on disjoint splitting must be abandoned, only that it cannot be the case that they eliminate the need for additional considerations to mark the boundaries of rational operators. More interestingly, along the lines discussed in section 3.2, localized trivial revision may be rationally justifiable in some circumstances. For example, suppose that one is trying to understand an economic theory that is contrary to one's own. In that case, it is perfectly reasonable to toss out all prior *economic* beliefs, but it would be (intuitively) beside the point to also toss out one's (presumably disjoint) beliefs about physics. The goal is to test out the wholesale replacement of one's economic beliefs, not of one's beliefs in general. Of course, if it turned out that one's economic beliefs were so intertwined with one's beliefs about physics that they were not separable, then, of course, there's a pretty strong sense in which the beliefs about physics *are*, also, beliefs about economics. What is somewhat compelling is that in the case where the topics are separable, and we want to throw out all beliefs concerning one topic, we would like to do so while preserving "all" beliefs on the other topics. (As we shall see in section 3.5, we must take care in the notion of "all".)

From the angle of human psychology, local trivial revision is nowhere nearly as uncanny as global trivial revision, at least. The idea that *everything* you believed about the mating habits of land snails is wrong is not particularly insane. One might doubt that it makes sense to throw out *all* one's beliefs about land snails *just because* it turns out that, contrary to prior belief, land snails are hermaphroditic. Usually, it does not. But if one's confidence in the belief of snail sexual dimorphism were very strong, finding out that one was wrong about *that* could give good epistemic ground for giving up the rest of one's snail beliefs. Or, the error might give good *methodological* grounds for giving up the rest *until* they are verified. The cases where giving up *all* one's beliefs upon getting new information are methodologically warranted (if even coherent) are pretty much limited to philosophical contexts and science fiction scenarios. Local trivial revision is, at least, occasionally plausible as a model of human belief revision where the belief sets are large, complex, and difficult to deal with.

Global trivial revision is more plausible in cases where either it has less significant consequences or the thing to be revised is more scoped. It is often perfectly sensible to delete an entire database upon discovering an error because the data can be regenerated from original sources, whereas trying to find and repair the error and all the effects of the error may be much, much more difficult and lead to unwanted side effects. We typically want to distinguish between the revision that eliminates an error due to a typo in a conversion script and a revision which expresses a disagreement with the original sources. It is not surprising that when we can, in principle, add back all the obliterated beliefs, global trivial revision is less frightening. One might argue that this is not really trivial revision at all but part of the internal dynamic play that constitutes the action of the revision operator. That is, it may be that what makes global trivial revision less frightening is that although the database passes through the wiped state, that state is not the *real* outcome of the revision.

There is some force to this concern, but not a huge amount. Since AGM theory is supposed to cover cases such as accepting a belief for the sake of argument, contrary to your own beliefs (see [Gärdenfors, 1988], page 48), it is hard to see why deleting all the sentences in a word processing document to replace them with a fresh start is out of bounds. In fact, people *do* irrevocably (or near irrevocably) throw away large works (the road through graduate school is littered with the corpses of abandoned theses), though they are more likely to radically revise when there is at least the perception of revoking the radical revision.

Given that the mere presence of trivial revision in the set of sanctioned operators is not enough to undermine a set of rationality constraints, and, in fact, given the stronger point that ruling out trivializing revision operators itself might be irrational, there does not seem to be anything left to Parikh's direct rationality argument for topic-relevance respecting operators. Global trivial revision can make sense for at least some, perhaps rare, cases. Local trivial revision is plausible, but that very plausibility undermines the brute horror of trivial revision in general. The standard AGM discussion which argues against trivial revision is best understood as an argument against trivial revision as the *only* sanctioned revision operator. Clearly, in most cases, trivial revision *is* too small, but that does not mean that it is *always* too small.

3.4 Computational Considerations

There is the hope that localized revision will be *easier*, at least in some circumstances. In particular, Parikh suggests that it will be computationally easier, as the revision operator will, in some cases, be applied to a smaller belief set.

In the case of large, collaboratively developed logical theories, localizing the change is a common and promising way of reducing both the communication overhead and the reasoning overhead.

There are several parameters we can vary when examining the computational properties of revision schemes including the consequence relation (i.e., the logic of the knowledge base), e.g., whether we are revising a belief base or a belief set; if we are revising a belief base whether revision is syntax oriented or model oriented (among other choices); in any case, we can vary the sorts of selection function used (or, relatedly, entrenchment relations). Furthermore, there are different services with regard to the revision that we might be interested in, for example, query answering against the revised beliefs, or computing a representation of the revised beliefs (e.g., [Liberatore, 2000].)

A point to notice straight off: splitting revision cannot have a lower worst case complexity than normal revision, since there are belief sets which cannot be split (e.g., if generated from a single literal or from a disjunction of positive literals). If the finest splitting of a belief set is the belief set itself, then applying the revision operator to the finest splitting is the same as applying it to the belief set.

Furthermore, there is the cost of finding the normalized base to consider. It is not a particularly easy task, and, in some cases (i.e., when the split base has only one component) it could be a complete waste of effort. Consider the propositional case with a finite vocabulary when we are given the belief set, B, by some finite base, B_b, which is in clausal form. In order to determine a split base for B, we make use of the set of *prime implicates* of B_b. An *implicate* of knowledge base is a clause (i.e., a disjunction of literals) entailed by that knowledge base. A *prime implicate* is a minimal length implicate, that is, an implicate such that if one deleted any literal from the clause, then the resulting clause would not be an implicate. Prime implicates have the interesting property that the set of prime implicates of a knowledge base entail the knowledge base. Quite often, the full set of prime implicates is redundant, that is, one can remove elements of the set and still entail the original knowledge base (in the boolean function lingo, one still has a *cover* of the original set). Some prime implicates may occur in every cover of the original knowledge base; these are the *essential* prime implicates.

Example 10. [8]

Consider the propositional knowledge base, B:

$$\{P \vee Q \vee \neg S, Q \vee S, \neg Q \vee \neg S\}$$

The prime implicates of B are:

$$\{P \vee Q, P \vee \neg S, \mathbf{Q} \vee \mathbf{S}, \neg\mathbf{Q} \vee \neg\mathbf{S}\}$$

(The bolded formulae are the essential prime implicates of B.)

There are two irredundant prime covers of B:

1. $\{P \vee Q, \mathbf{Q} \vee \mathbf{S}, \neg\mathbf{Q} \vee \neg\mathbf{S}\}$*; split by* $\{\{P, Q, S\}\}$

[8] This is example is taken from one in [Hammer and Kogan, 1996].

2. $\{P \vee \neg S, \mathbf{Q} \vee \mathbf{S}, \neg \mathbf{Q} \vee \neg \mathbf{S}\}$*; split by* $\{\{P, Q, S\}\}$

(Obviously, each cover must contain the essential prime implicates.)

Now let us consider $B' = B \cup \{S\}$*. The prime implicates of* B' *are:*

$\{P, \neg Q, S\}$

Which is also the only minimal cover. It is also a split base, where the splitting language is $\{\{P\}, \{Q\}, \{S\}\}$.

Prime implicates are used for compilation of knowledge bases, for example, to produce propositional theories for which query answering is in P (arbitrary propositional query answering is co-NP-complete). Of course, this "merely" shifts the computational pain to the compilation step and perhaps involves an exponential increase in the size of the knowledge base [Reiter and de Kleer, 1987]. Knowledge compilation is somewhat analogous to Kourousias and Makinson's preprocessing approach, except that in knowledge compilation, typically the same operation is performed on the original and on the compiled knowledge base. Furthermore, the operation should be strictly less complex in the compiled setting (or the compilation is pointless) [Liberatore, 2001]. We can use prime implicates to compute a finest split base as in [Bienvenu et al., 2008].

It would be a happy coincidence if the maximal split base were also a compilation. For some revision operators and forms of theory this is true.[9]

[9] Though this is a complex topic: for example, instead of compiling to a form for which your service is tractable in the size of the compiled form, but the compiled form may be exponential in the size of the original form, one might reasonably choose some form which is not as tractable, but which is much smaller than the original.

Some attention to splitting the given base is also reasonable: by the finest splitting theorem, we know that *every* splitting is an approximation of the finest one. So, at the very least, for a given input, one can reduce the amount of compilation required by first splitting the base, then getting the implicates of only the partitions relevant to the input. Splitting a given base is very cheap as it is basically a sort of the signatures of the axioms. Given base splitting is potentially helpful either for base revision or belief set revision, and one can achieve better approximations by performing obvious transformations on the base, e.g., splitting conjunctions. Given base splitting may be stymied by semantically inessential connections, as seen in example 9.

But in the end, the computational advantages of splitting, while often promising, seem more ambiguous and less profound than we might have hoped. It is true that "dividing and conquering" is a useful strategy, in general, but that aspect of splitting just seems to be one among several (compare with compilation or syntax sensitive base revision). At the very least, it depends strongly on the logic in question and whether computing the smaller portion is easier than computing the revision against the entire knowledge base. In many cases the complexity may coincide exactly.

3.5 Naturalness

We come to the final justification for Parikh-relevance respecting revision: its "naturalness". This is distinguishable from Parikh's anti-trivial revision argument essentially in that the naturalness argument can be used to justify globally trivial revision operators *when* they respect Parikh-relevance, e.g., if the finest splitting has only one element. It also can provide support for the rationality of local trivial revision operators, whereas it is hard to ground endorsement of local trivial revision on revulsion for trivial revision in general. If Parikh-relevance can be justified on independent grounds, then we do have a basis for distinguishing between local and global trivial revision, that is, we are not forced by the principle of informational economy to prefer operators that presume more beliefs (overall) to those which prefer more topic-irrelevant beliefs.

Topic orientation *is* very natural. Intuitively, changes with regard to one's belief about a certain subject matter should be independent of disjoint subject matters, that is, of subject matters that the input does not touch upon. This consideration smells a bit of foundationalism, especially when phrased in terms of relevance. But topic orientation does not necessarily prejudice the case against coherentism [10]. While the fact of unique finest splittings suggests that the associated splitting formulae, that is, the equivalence class of normalized bases *is* the foundation, one could argue than any properly coherent belief set will have a singleton splitting. Imagine the merge of two belief sets with disjoint signatures. It seems rather extreme that their mere set theoretical union should generate strong, epistemically significant connections between the two domains. Instead, a coherentist could ask that coherence itself be non-trivial. In this scenario, non-singleton splittings are a *measure* of incoherence. Such a scenario is fairly realistic too. While one *could* treat the information as uniform, it would be more sensible to organize it into parts

[10] For discussions of interpreting base revision as modeling justificatory foundationalism and of belief set revision as modeling justificatory holism or coherentism, see [Gärdenfors, 1990, Doyle, 1992, Bochman, 1999]. Also, [Bochman, 1998] gives an account where standard AGM theory can be seen to support a foundationalist approach.

that can stand alone, where everything in stand-alone parts is required, and and those parts capture the rest of the information. In other words, coherentists do not have to be against there being significance to logical structure.

The fact that there is a univocal finest splitting strongly buttresses the naturalness of Parikh-relevance. The finest splitting is a striking feature of the *logical structure* of a belief set that we would seem to ignore at our peril.

Clearly, there are cases where strict topic orientation is what we require. Collaborative authorship of textbooks, for example, often split along topic lines *when there are strictly disjoint topics to be found*. Even where there are not, the need for *independent* evolution of a the book leads us to treat subparts as effectively disjoint (or, at least, separable).

Similarly, large biomedical or bioinformatics ontologies tend to be collaboratively developed with spatially and temporally distributed teams of editors. Cognitively, it is much easier for developers to focus on logically coherent subsets of the whole. The difference between having to work with tens of thousands of classes and only thousands can be essential to the effectiveness of various cognitive support techniques, such as visualization.

However, topic-orientation has some potentially counterintuitive results, as articulated by Kourousias and Makinson:

> [Suppose] we are contracting the letter p from the closed belief set $K = \mathrm{Cn}(p, q)$, or from its base $K_0 = \{p \leftrightarrow q, q\}$. If we are working with the base, we may perhaps regard it as supplying us with epistemic information,

> namely that its elements are particularly important items that deserve to be protected more than other items not appearing in it. But even so, the base gives us no information to discriminate *between* its elements. In the example $K_0 = \{p \leftrightarrow q, q\}$, we are not told explicitly that the element letter q deserves protection more than the biconditional $p \leftrightarrow q$. Indeed, we may wish to allow for the possibility that the latter is more deeply entrenched, less vulnerable, than the former. In that case, when we discard p we will jettison the letter q and keep the biconditional, regardless of the fact that q is irrelevant to p modulo K in the sense that Parikh has defined.
>
> On the other hand, there may be occasions in which we wish to treat elementary letters systematically as the only carriers of epistemic significance. In our view, this policy is difficult to justify in theoretical terms, but it sometimes appears to be adopted in contexts of artificial intelligence for reasons of computational convenience. Under such a policy, compound formulae such as the biconditional $p \leftrightarrow q$ are of no epistemic significance, and when changing a belief set we would want to minimize change in the status of elementary letters and preserve relevance in the sense of Parikh.

The first point to notice is that the alleged bias toward sentence letters ("elementary letters [are] the only carriers of epistemic significance"!) is not a bias towards sentence letters (i.e., literals) per se. Consider the a variant of the belief set K, $K' = \mathrm{Cn}(p \leftrightarrow (q \vee r), q \vee r)$, where we replace q with the disjunction $q \vee r$. All of the above considerations go through, except that we do not have any literals to preserve. Parikh-relevance is biased toward formulae which have *smaller signatures*, and literals obviously benefit from that. Furthermore, this bias is not motivated primarily, by Parikh, by computational considerations (though computational considerations come up), but by the attempt to subdivide the signature as much as possible. It is fair to dispute that topic separability is not a rationally overriding value, but it is just misguided to assert that there are no independent theoretical considerations in play.

One strange aspect of the discussion is that Kourousias and Makinson are not defending, or even discussing, untoward global trivial revision (e.g., full meet contraction based revision). (Strikingly, though trivial revision is the expressed motivation for Parikh, they do not discuss it *at all*.) We could ask, "Is AGM's failure to rule out global trivial revision really a shortcoming?" but as the standard accounts reject full meet revision (and some associated proposed axioms) except as a limiting case, this does not change matters. Alternatively, we might ask "Is ruling out global trivializing revision when it violates Parikh-relevance worth changing AGM theory?" To answer this, we need to know what we're giving up. Kourousias and Makinson claim that we would have to impose an epistemic significance scheme that at best has only a pragmatic, computational "convenience" justification (i.e., that atomic sentences are most important). Since there is, to their eyes, no *general* logical consideration, nor, indeed, a generalizable pragmatic consideration (such as, the revised set being "too small"), there is no reason to make Parikh-relevance a basic criterion of rational contraction and revision. As one sort of consideration, Parikh-relevance may be a reasonable filter on revision operators. That is, one epistemic value that an agent might hold is topicality, but this value is not forced upon us. In their example, it is easy to imagine that $p \leftrightarrow q$ was vouched for by a less dodgy source (or was simply older) than q, and so, especially in a base revision context, should be preserved.

However, in a base revision context, we might well take the finest splitting as the entire base (K_0) since *those sentences* have overlapping signatures (q appears in both), and thus cannot be separated. Kourousias and Makinson slide a bit too quickly from considering the base, as such, to considering general Parikh-relevance. Traditionally, there is quite a bit of difference between belief set oriented revision and base oriented revision. If we take K_0 as a given base (rather than a generating base, i.e., a base which is merely a description of the corresponding belief set), then it is unclear why we would apply standard, belief set oriented AGM principles, even as modified by Parikh. So, it is not clear that q *is* irrelevant to p. Indeed, the example strongly suggests otherwise.

If we retreat to the corresponding belief set, K, one can still frame Kourousias and Makinson's worry in terms of epistemic entrenchment. We can certainly formally associate an entrenchment relation over K such that the biconditional is more entrenched than q, but it is a bit trickier to assess the plausibility of such a relation. If the relation is derived from something like provenance of the belief, then we start to slide back into a belief base situation.

On the other hand, revising K by $\neg o$ to get $K_1 = \mathrm{Cn}(\neg p, q)$ vs. $K_2 = \mathrm{Cn}(\neg p, \neg q)$ can be seen as turning on whether one believes the conditional $\neg p > \neg q$ or not (or, really, the biconditional, but only the one direction matters for this example)[11]. Obviously, by the Ramsey test, if one does endorse that conditional, then one should revise to K_2 (see the discussion in chapter of [Gärdenfors, 1988]) and thus violate Parikh-relevance.

[11] There are well known issues with the complete set of AGM postulates and the semantics of conditionals, see section 7.4 of [Gärdenfors, 1988].

The epistemic significance priority of $p \leftrightarrow q$ is a way of encoding that conditional (or, conversely, the conditional encodes the priority).

If the entrenchment relation is supposed to ground a conditional belief of the form $\neg p > \neg q$, then we should *consider that conditional* when determining the topic structure of our belief set. $\neg p > \neg q$ clearly connects p and q in in a very strong way. If we added this conditional to K as an explicit belief, it is hard to see how the adjusted notion of Parikh-relevance would not respect cross-topic conditionals.

In essence, we're back to the coherentist question of how seriously to take arbitrary material conditionals. In classic AGM theory, there was no sense that belief sets had any structure other than what was imposed upon them by entrenchment (which is, after all, an extralogical consideration). Base revisionists rebel against that but again impose an extra logical condition (i.e., the base is just a brute given). Splitting gives us an inherent, logical reason to neglect certain sentences in a belief set and a reason to prefer certain classes of base. Of course, different normalized bases might be preferred for any number of non-logical reasons and we might even choose to avoid normalized bases in some circumstances. Similarly, we might override splittings by an entrenchment relation. But these are, or should be, departures from the norm. These *override* what the logic of the belief set tells us. We might have good reasons for that overriding, but we cannot claim that it is not an overriding.

3.6 Splitting and Transitively Relational Partial Meet Contraction

By their conformance to the core AGM postulates, Parikh-relevance respecting contraction operators are also partial meet contraction operators. If they are, as Parikh suggests, incompatible with the supplementary axioms, then they cannot correspond to relational and transitively relational partial meet contraction operators. That is, it would be the case that selection functions defining Parikh-relevance respecting contraction operators cannot be defined in terms of the most entrenched elements of the remainders. However, there are cases where Parikh-relevance is not incompatible with transitively relational partial meet contraction, with the obvious case being when there is only one element in the finest splitting.

It is certainly possible to build a reasonable story around this degenerate case. The first point is to grant that *if* one is going to use remainders (i.e., maximal subsets which do not entail the to-be-contracted formulae) as a fundamental tool for defining contraction, and thus revision, then using an ordering (and a transitive ordering) to define the selection function is overwhelmingly compelling. However, it is unclear that remainders are a particularly natural object of epistemic focus, at least in the sense that one is obliged to base one's extra-logical justificatory structure on them. Compare with safe contractions [Gärdenfors, 1988], where there is a ranking between formulae in the belief set. As Hans Rott [Rott, 1992] writes:

> [S]afe contraction by its very idea focusses on *minimal sets of premises* sufficient to derive a certain sentence. Thus safe contraction has a certain "foundationalist" appearance, in contrast to the "coherentist" guise of its

> competitors...
>
> Safe contractions appear to posses some definite epistemological advantages over both partial meet and epistemic entrenchment contractions. Like epistemic entrenchment contractions, they are based on some kind of relation between *sentences* and not on a relation between *sets of sentences*, as it is the case with partial meet contractions. This constitutes an intuitive disadvantage of the latter. In addition, safe contractions rest on relatively weak requirements for the relation involved, which seem to give them the intuitive priority over epistemic entrenchment contractions.

Unfortunately, the intuitive and epistemological appeal of safe contraction over the alternatives does not provide us with grounds for rejecting the supplementary postulates. In the original safe contraction paper [Alchourrón and Makinson, 1985], Alchourrón and Makinson show (for the case of finite belief sets) that suitable safe contraction operators are also transitively relational partial meet contraction operators (and vice versa). As pointed out in [Gärdenfors, 1988], the suitable safe contraction operators are transitively relational partial meet contraction operators even in the infinite case — a much worse fact from the current perspective.

The notion that partial meet contraction is especially "coherentist" does open up a rather different line of attack. If a belief set can be split into more than one topic, then the natural reading, as we have argued, is that the belief set is not particularly coherent. That is, the separability of topic indicates that some of the "beliefs" in the belief set are not strongly mutually supporting. Indeed, they *do not talk about the same things*. Thus, we can take the different levels of coherence within the belief set as *indicators* as to when coherentist style considerations *should* apply. That is, coherentist style operators should apply only to strongly *logically* coherent fragments of a belief set. If the belief set as a whole is strongly coherent (i.e., there is only one element to the finest splitting) then coherentist considerations take over altogether. Similarly, within a logically coherent fragment, we can treat things in the standard AGM way, but outside it we have to mobilize other considerations. This thought is almost a paraphrase of Axiom P.

While appealing, this approach will not fly. The problem is that a new formula can make the belief set more *or* less logically coherent. This is easily seen:

Example 11. *Let* $K = \mathrm{Cn}(P, Q)$ *and consider what happens if we revise by* $\neg P \wedge \neg Q$ *(respecting Parikh-relevance). In this case the signature of the input formula forces together all the elements of the splitting (as is appropriate!).*

Consider revising $K = \mathrm{Cn}(P \vee \neg Q, Q)$ *with* $P \vee Q$ *(which reduces to expanding by* $P \vee Q$*).* K *is fully coherent; the input signature equals* K*'s signature, but the resulting belief base is split by* $\{\{P\}, \{Q\}\}$.

Part of the difficulty is that remainders and splittings cut across each other, making them difficult to reconcile. But perhaps Parikh was too quick to dismiss the possibility of compatibility between Parikh-relevance and the supplementary postulates. After all, there are partial meet and safe contractions which do not satisfy the supplementary postulates! In fact, we can show that it is not particularly difficult to generate a (safe) contraction operator which respects Parikh-relevance. Recall that a safe contraction operator requires[12] an acyclic relation, $<$, over the formulae in a belief set called a *hierarchy*. To conform

[12]In this section, we shall discuss only the aspects of a safe construction operator needed to show con-

with C7, it is sufficient that $<$ *continues up*[13] the consequence relation. $<$ continues up $\models$ iff $\models$ transmits $<$. That is, if a formula is $<$ some other formula, it is also $<$ all the consequences of the second formula. To conform with C8 as well, it is sufficient that $<$ is (additionally) *virtually connected*. That is, "for all A, B, C in K, if $A < B$, then either $A < C$ or $C < B$"[14]

Given these facts, it is easy to show:

Theorem. *There are Parikh-relevance respecting contraction operators that conform to C7 and C8*

Let K_n be the normalized base of a belief set K.

We define a hierarchy, $<$ over K that meets the following conditions:

- *for all $f \in K_n$ and $g \in K \setminus K_n$, $g < f$,*
- *for all f_i, f_j $(i \neq j)$, $f_i \nless f_j$*
- *otherwise, $<$ is virtually connected and continues up over the consequence relation of K*

It is easy to see that $<$ is virtually connected and continues up as a whole. Since there are no formulae, x such that an element of $K_n < x$, the formulae in K_n vacuously meet continuing up. Similarly, for virtual connectedness.

The elements of K_n are safe *with regard to some formula C unless there is a subset,*

formance with the supplementary postulates.

[13]Or continues down, but we ignore that for now

[14][Gärdenfors, 1988], pg. 98. Note that the first condition corresponds to Theorem 4.34 of [Gärdenfors, 1988], and joint conditions to theorem 4.35.

K_n', *of* K_n *(and thus a minimal subset) which entails* C*. By Craig's interpolation,* $\mathrm{Sig}(K_n') \cap \mathrm{Sig}(C) \neq \emptyset$*, which is exactly what's needed for Parikh-relevance.*

Thus, there is no strong incompatibility between Parikh-relevance respecting operators and transitively-relational partial meet ones. So, Parikh-relevance cannot be used as an argument against C7 and C8, at least, in general.

What are we to make, then, of Parikh's motivating example? One very strange feature of it is that it actually violates a *core* AGM postulate!

Example 12. *Recall in Parikh's example,* $A = P \vee Q$ *and* $B = \neg P \vee Q$*. His point was that revising by* $A \wedge B$ *could preserve more* P *beliefs than revising first by* A *then expanding by* B *(for example, in the case of local trivial revision), since* $A \wedge B$ *was equivalent to* Q *alone "and says nothing about* P*".*

But by extensionality, revisions by equivalent formulae should be equal, which is not the case in this example:

Let $C = (P \vee Q) \wedge (\neg P \vee Q)$ *(i.e., the conjunction of* A *and* B*) and* $D = Q$*.*

Clearly, $C \leftrightarrow D$*.*

Now, suppose we have a belief set, $K = \mathrm{Cn}(P, \neg Q)$*. Using the local trivializing revision operator, we get:*

$$K * C = \mathrm{Cn}(Q) \; K * D = \mathrm{Cn}(P, Q)$$

Thus, $K * C \not\leftrightarrow K * D$*.*

Violating extensionality is a far more serious issue than violating the supplementary postulates. In classic AGM theory, there are no acceptable operators which violate a core axiom. Furthermore, this example violates the principle of irrelevance of syntax in a fairly egregious way: syntax is irrelevant in the belief set but not in the input, and it is only relevant in the input for as long as it is considered *as* input.

Fortunately, the remedy is fairly straightforward: normalize the input as well. If we take the supplementary postulates to apply to normalized inputs, then Parikh's counterexample fails, since Q is not the conjunction of A and B, even though it is equivalent to that conjunction. But now there's no reason to be concerned about the fate of postulates 7 and 8.

3.7 Splitting in other logics

Thus far, our discussion has been restricted to propositional logic. Since one of the strongest arguments for Parikh-relevance has been the univocality of the finest splitting, it is important to see how widely it extends.

Parikh claims that finest splitting theorem holds for predicate logic without equality (and fails in the presence of equality). This can be seen more easily if one considers that proofs of the finest splitting theorem given by Kourousias and Makinson. They show the finest splitting theorem by appealing to Craig's interpolation (and to their parallel interpolation, but that is entailed by Craig's interpolation). Thus, for logics with Craig's interpolation, we can prove that there is a univocal finest splitting. In logics where interpolation fails, such as predicate logic with equality, it is not, therefore, surprising that univocal finest splitting fails. In predicate logic with equality, the failure can be easily seen more directly, as Parikh points out: equality can be used to fix the size of the interpretation domain to a given finite size. In such cases, formulae with disjoint signatures may interact by contending for elements of the domain without any mediating signature.

The interpolation property may fail for other reasons, including not having enough syntactic richness to express interpolants for all cases (not uncommon in modal and description logics). In (some of) these cases, unlike the equality case, we can point to a finest splitting that is inexpressible in the logic in question, but expressible in a richer logic. For example, if the logic in question is embeddable in predicate logics without equality (as many modal logics are), we can appeal to the splitting of the belief set in question with regard to its predicate logic translation. Since belief bases in this case will also be univocal and approximations of a univocal finest splitting (inexpressible in the current logics), we have a strong reason to focus on base revision.

3.8 Conclusion

In this chapter, we examined Parikh's account of topic-oriented belief revision based on *splitting* the belief set according to disjoint signatures. We showed that Parikh's most prominent argument for it (blocking trivial revision) fails to compel, but that the major arguments against it (incompatibility with the supplementary postulates; untoward priority to literals) also fail. Moreover, the *naturalness* argument, at least for many areas, seems quite compelling.

In particular, we find the existence of an identifiable (indeed, computable), univocal logical structure to be impossible to ignore. While one can imagine cases where one might wish to override that structure, many of those cases are subsumed by a shift to belief bases, and, indeed, to syntax orientation. There, while the conditions of the structure have shifted, the imperative to take that structure into account remains.

Chapter 4

Coping with Inconsistency

In chapter 3, we developed a Parikhian argument to the conclusion that there is a class of rational (even AGM rational) belief revision operators which respect a form of topic oriented relevance. A belief set (or base) is divided into coherent parts such that each part has a signature disjoint from that of every other part, each signature is maximally small with respect to the belief set in question, and the characterizing formulae of each topic jointly entail the original belief set. Topic relevant revision (or contraction) thus begins with identifying the topics of the input formula (i.e., the set of topics which intersect with the signature of the input). Topicality, in this sense, has some intuitively pleasing features: Topicality is univocal and derived solely from the logical structure of the belief set. It also allows for a more natural coherentism: beliefs don't cohere with *every* other belief, only with topic relevant ones. Revision thus does not focus on what is narrowly relevant (or implausibly distant), but on the proper subject matter at hand. Topics are a kind of modular decomposition of the belief set, and revision which respects topicality gains some of the advantages of modularity, e.g., encapsulation of information and isolation of effects. This may have computational benefits (assuming reasonable amortization of the cost of determining the topic structure) and definitely has cognitive benefits.

When changing belief sets, we *revise*, rather than expand, the set when there is a conflict between our belief set and the new information. The conflict traditionally of concern in the belief revision tradition is contradiction, that is, the belief set contains the negation of the new formula. Parikh-relevance sensitive revision *isolates* the possible conflict to the relevant topics. But if we can isolate the *possible* conflict, then it seems similarly possible to isolate *extant* conflicts, that is, to isolate contradictions in the object of revision. Clearly, the object of revision containing topic-isolated contradictions cannot be a traditional belief set with a supra-classical consequence relation since all such belief sets would be the same (i.e., the set of all formulae of the language). One prominent strand of paraconsistent belief revision focuses on belief bases for their ability to distinguish between different contradictory sets of beliefs. For example, while $\mathrm{Cn}(A, \neg A, B)$ and $\mathrm{Cn}(B, \neg B, A)$ are the same (for, e.g., propositional logic), the generating bases are obviously quite different. One obvious way to notice the difference is that they can be seen as the result of expanding the same belief set $\mathrm{Cn}(A, B)$ with the distinct, nonequivalent beliefs $\neg A$ and $\neg B$, respectively. Of course, the actual resultant belief set is the same in both cases, but that's because we have an extra step of generating the deductive closure. If we focus on the *process* (rather than the result), it is clear that they are very different circumstances. If we replace the expansion with a revision, that difference is clear. $\mathrm{Cn}(A, B) * \neg A$ is very different from $\mathrm{Cn}(A, B) * \neg B$, especially if the operator a locally trivializing topic sensitive one (i.e., in the first case we get $\mathrm{Cn}(B)$ and in the second $\mathrm{Cn}(A)$). It is an unfortunate and surprising feature of belief sets[1] that expansion and revision in the

[1]This is especially clear if one considers not just models of a collection of beliefs but interpretations. Because belief sets are consequence oriented (and thus, strangely enough, *syntactically* oriented) they cannot distinguish between inconsistent theories with different sanctioned sets of interpretations. For example,

inconsistent case produce such radically different results.

Semi-revision, as described in [Hansson, 1999a] is also sensitive to the difference between these revisions. Semi-revision is an operation on belief bases and is defined in terms of expansion and *consolidation*, that is, an operation that takes in an inconsistent belief base and returns a consistent one. Semi-revision is expansion by the (contradictory) input, then consolidation of the inconsistent expanded belief base. Since the input formula is not privileged by the consolidation operator *because of its input status*, semi-revision, thus defined, is non-prioritized.

Semi-revision requires being able to distinguish various inconsistent belief states (thus, standard belief sets are not feasible representations), but semi-revision operators consume consistent belief bases and produce consistent belief bases. Inconsistent belief bases appear only as transitional states. This reasonably models the key distinction between standard AGM revision and semi-revision: semi-revision doesn't require success. Since the inputs and outputs of semi-revision are consistent (and possibly normalized) bases, it does not seem especially difficult to impose some sort of relevance criterion. For example, a simple approach would be to normalize the initial base; extract the parts of the normalized base whose joint signature is the minimal joint signature that is a superset of the signature of the input; expand the extraction with the (conflicting) input while leaving the rest of the normalized base untouched; finally, we can consolidate the (now

intuitively, if you take a theory $\mathrm{Cn}(A, B)$ then clearly all the models will assign true to A and false to B. If we add $\neg B$ then it's pretty clear that we'll want to consider interpretations where B is assigned true and ones where it is assigned false (and thus $\neg B$ assigned true). But there's no reason to consider interpretations that assign false to A.

contradictory) extract part independently of the rest.

At this point, the fact that we can isolate, at least in the dynamic context, the contradiction to its relevant (set of) component(s) of the splitting of the original theory (since this is what sets up the target of consolidation) strongly suggests that we can split, and perhaps normalize, a contradictory belief base. Parikh (with collaborators) went in this direction, so we shall investigate their takes first.

4.1 Motivations

In [Chopra and Parikh, 1999] Samir Chopra and Rohit Parikh make the following three (naive) observations about people's doxastic architecture:

1. People often reason in a topic focused way; that is, with a non-arbitrary subset of their beliefs;

2. People often have inconsistent beliefs;

3. People solve what are computationally intractable problems, e.g., propositional reasoning.

The first and third observations are familiar from [Parikh, 1999] as motivations for Axiom P and as motivations for selecting topic sensitive revision operators, are extensively discussed in chapter 3. One point worth noticing is the much stronger psychologistic flavors of the considerations in [Chopra and Parikh, 1999]. For example:

> 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 belief". If computational complexity were a problem in real life, such answers would be common-place.

It is tempting to treat this comment (as well as the other motivations derived from naive psychology) as boilerplate introductory throwaways. However, Chopra and Parikh treat these considerations as evaluative criteria for a theory of belief revision. Indeed, they claim that their model of belief provides " psychologically plausible, computationally tractable procedures for belief revision". This is no mere buttressing consideration, nor is it touted as a mere pragmatic advantage of their theory for the sake of engineering human -ike systems, but as part of the basic constraints of an account of revision rationality.

It is still tempting to treat these comments as throwaways because of their extreme silliness. The first question that comes to mind is asking *what it is* that human beings are not deterred from doing by the co-NP-completeness of propositional derivation. How many people even know what propositional derivation is, much less its complexity, much less the significance of that complexity? (We've encountered many graduate students in computer science who aren't clear on even the common cant about the significance of intractable problems, much less the many subtleties of how computational complexity considerations should affect practice.) It seems much more likely that sheer ignorance is a better explanation of lack of deterrence. Furthermore, not many people try to derive things in propositional (or any other) logic outside very specialized contexts. So even if they would be deterred from trying to find propositional derivations, that's hardly an interesting consideration. (The intractability of 3SAT doesn't even deter people from writing propositional satisfiability solvers!)

Finally, it is very hard to see why anyone thought that, for example, AGM with propositional belief sets was psychologically realistic, that is, was anything remotely resembling what goes on inside our heads. Even if it were a reasonable approximation, it is very hard to see that people in normal interaction would characteristically distinguish between knowing something and knowing whether it follows from what they explicitly believe. In fact, people do so distinguish as the disputation tactics of *reductio ad absurdum* and (non-fallacious) *ad hominem* illustrate. It is very common to acknowledge that there are consequences of our beliefs that we are unaware of.

This sort of problematic psychologism is not uncommon in knowledge representation. For example, Franz Baader, in [Baader, 1999] writes:

> Another requirement that is usually imposed on KR formalisms is that of allowing for a *structured representation* of the knowledge. This means that semantically related information (for example, all the knowledge about knowledge representation based on Description Logics) should also syntactically be grouped together. This requirement is, on the one hand, justified by cognitive adequacy. [Footnote: In the human brain, correlated information is also not stored in unrelated parts.] On the other hand, there are purely pragmatic reasons, since structured representation allows for faster retrieval.

Gerhard Strube (in [Strube, 1992]) identifies two concepts of cognitive adequacy, strong and weak. A formalism or representation exhibits strong cognitive adequacy if it correctly models human concepts or cognitive processes. A formalism or representation is weakly cognitively adequate if it is *workable* for human beings. The former is a claim about suitability as an account of cognition, whereas the latter is a claim about usability. These two claims do not necessarily go together. It may be the case that a formalism is not particularly usable due to its strong cognitive adequacy: for example, if the phenomenon itself is complex and difficult to understand or contains many fine details that are irrelevant to some task at hand, then it is quite likely that the model of it will be complex and hard to work with. Similarly, if human thinkers are quite bad at some task, an accurate model is not promising as a candidate prosthesis.

Baader's appeal is fairly superficial and buttressed by pragmatic performance considerations (which, in fact, are equally superficial in the way that Chopra and Parikh's are). One can tie performance, in part, to weak cognitive adequacy as responsiveness is an important consideration for usability. Also, there is a hint that it is not accidental that strong cognitive adequacy should indicate other benefits: after all, we are the most successful knowledge based systems we know of. The flip side is that the presence of the ancillary benefits is a sign that our formalism is strongly cognitively adequate.

These last two lines are rather worrisome. Our understanding of how humans think, and our command of "intelligent" technology, is so rudimentary, we should be very wary of suggestions that the way forward is to replicate "how we do it". Consider how difficult a task it would be to replicate the human intestine, or even a kidney, where we have something close to an equivalent (for some of the functionality). Or consider thinking that we had to closely mimic a kidney before making a dialysis machine. Or insisting that mechanized locomotion devices had to be bipedal. Existing systems are, of course, worth studying, drawing inspiration from, and even (crudely) emulating. But it is not fruitful to constrain the design of useful, usable tools by an emulation criterion.

Thus, if our concern is usability, strong cognitive adequacy, in itself, is not interesting. Traditional user studies will be far more relevant than the kind of experiments described in [Renz et al., 2000]. Interestingly, often there is no need to ratchet up the adequacy stakes. A plausible *prima facie* case for even the simple kind of structuring Baadar mentions can be made by observing our general, external information management practices. After all, librarians have long held that semantically related information should also be syntactically grouped together. A library doesn't cognize, but such organization does allow for faster retrieval.

Similarly, people do things in topic oriented ways. They think, write books, talk, organize libraries, categorize email, diagram arguments, and build ontologies with quite a bit of regard to topicality. Indeed, the very task of determining what is topically related to what is a major activity! For some of these activities, computational complexity is not a particularly interesting question. For some, for example, ontology development, computational complexity is very important as we want to automate some tasks.

So, in spite of their problematic origin, these considerations are useful both as input to rationality determination (as seen in section 3.5 and earlier in this chapter) and pragmatically. We can accept their spirit even if their letter is wanting.

Technically, the key issue is whether we can extend topic oriented revision to the inconsistent case. As we shall see, the account in [Chopra and Parikh, 1999] falls short on some critical points.

4.2 *B*-structures

[Chopra and Parikh, 1999] propose a model of belief representation and revision that attends to the following four concerns (stripped of their psychological flavor):

1. Belief revision should meet Gärdenfors' preservation criterion [Gärdenfors, 1988], that is, that a revision operator should retain as much of the original theory as possible.

2. Operations on belief representations need to be computationally practical.

3. A belief representation system should deal sensibly with inconsistent beliefs.

4. Belief representations should allow for some portion of the belief state to be implicit in the representation.

The first two concerns were met in [Parikh, 1999], which proposed adding an axiom to the standard core six axioms of AGM theory. However, as stated, AGM theory amended with Axiom P (what they call "LS" theory) does not handle the remaining two concerns: inconsistency tolerance and a robust distinction between implicit and explicit beliefs. It is unclear if there is a robust, independent motivation for this last concern other than inconsistency tolerance.

In order to handle the remaining concerns, Chopra and Parikh introduce the notion of a belief structure (*B*-structure) which deviates from Parikh's extension of AGM theory (LS) in three ways:

1. They shift from the traditional AGM focus on logically closed theories to belief sets

that may be not closed (e.g., the "object of revision" may be a set of beliefs which does not explicitly contain all the consequences of the set)[2],

2. they relax the disjointness requirement on splittings, and

3. they allow for the theory associated with the B-structure to be inconsistent before *and after* revision.

Parikh's splitting languages are strictly disjoint sets of atomic terms. Instead of a strict partitioning of the language L (and thus of the signature of formulae in the normalized base), Chopra and Parikh allow overlap between the sublanguages. Instead of a splitting, we have a B-structure:

Definition 6. *([Chopra and Parikh, 1999]) For a set of primitive terms, L, let* $\{L_1, \ldots, L_n\}$ *be such that* $L = \cup L_i : i \leq n$ *and* $\{T_1, \ldots, T_n\}$ *such that for all i,* T_i *is a theory in* L_i*.* $\{(L_1, T_1), \ldots, (L_n, T_n)\}$ *is a B-structure on L.*

When $n = 1$, then we have a normal theory T in L. When $\{L_1, \ldots, L_n\}$ are disjoint (that is, strictly partition L) and $T = \mathrm{Cn}(\{T_1, \ldots, T_n\})$, then $\{L_1, \ldots, L_n\}$ is a T-splitting. So all splittings have related B-structures, but, due to possible signature overlap, not all B-structures are splittings. Chopra and Parikh motivate allowing overlap by mobilizing intuitive considerations against the disjointness of (logical) topics:

> 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.

[2] A standard move, e.g., see chapter 3 of [Hansson, 1999b].

> 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.

Unfortunately, this example is problematic. As we saw, split theories do contain sentences with signatures that straddle more than one element of the splitting; indeed, it is impossible for them to not have such. Just consider tautologies over the whole signatures. Secondly, it's not clear that this is a case of *topics* being fluid in a static structure. If we have a belief set wherein Clinton and Lewinsky are part of a splitting element that is disjoint from the element concerning Iraq, then, when we read in the newspaper that Clinton is the one who is bombing Iraq we have a *change* to our belief set. This change alters the topic structure of our belief set. But this should not be surprising at all. Clearly different theories are going to have different logical structure.

It may be the case that these topics, in any appropriate regimentation, are *never* going to be (logically) separable into disjoint areas. Clinton is a person, as is Lewinsky, as are the bomber pilots, and the Air India pilots. It just might be the case that common sense is highly non-modular and thus the finest splitting of a *faithful* representation always will be identical with the whole regimentation (and thus too bad for Parikh-relevance in that context). However, there are other contexts where things are more structured.

In any case, a B-structure, $\{L_1, \ldots, L_n\}$ does not have to strictly partition L, instead it may k-partition L:

Definition 7. *[Chopra and Parikh, 1999] Let* $\{L_1, \ldots, L_n\}$ *be such that* $L = \cup L_i : i \leq n$.

$\{L_1, \ldots, L_n\}$ *k-partitions L if any symbol in L appears in at most k elements of* $\{L_1, \ldots, L_n\}$.

Finally, B-structures may be inconsistent to varying degrees: in particular, while each element of the B-structure must be consistent, sets of elements (due to the sharing of symbols) may be jointly inconsistent.

Definition 8. *[Chopra and Parikh, 1999] A B-structure is m-consistent if there is no set of* T_i *of cardinality m such that the* T_i *are jointly inconsistent.*

Chopra and Parikh require that each component, T_i be consistent, that is, that all B-structures are 1-consistent (and thus all inconsistent B-structures are at least 2-partitioned). The discussion of this requirement is a bit thin:

> 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 . . . Even a single agent maybe thought of as consisting of a collective entity. . .

The second paragraph aims to provide an independent motivation for the local consistency criterion, that is, that it is helpful for modeling groups. But this is independent as it requires a justification for the consistency requirement on the individual agents (in the group) in the first place. Presumably, these "agents" can also be human beings and all the psychologistic points hold (i.e., such agents do have inconsistent beliefs and can function in group situations; they can, and do, have both intra- and inter-agent conflict). Indeed, the first paragraph claims that most agents are in fact globally inconsistent! So how is it that in the second paragraph we can impose an agent-global consistency constraint?

It is perhaps technically convenient to reuse classical reasoning at least at the local level (much in the way that Kourousias and Makinson reuse the standard AGM axioms on a normalized base), but it is hard to see why it is more than a technical convenience[3] in this case. Even if we embrace a homuncular view of individual agents, it is unclear

[3]Or a dialectical convenience, albeit a dodgy one. There is sometimes the feeling that if one can somehow claim not to have modified classical logic, but just to be using it in a clever way, that this is somehow inherently less suspect. In reply, we appeal to [Makinson, 1982]:

> The authors are at pains to insist that their approach is not based on a non-classical logic, saying for example that "our logic can continue to be altogether classical" (p. 22) and "the present approach . . . dispenses entirely with any need to modify the principles of classical logic" (p. 58). In one respect this is perfectly true: the definition of an intelligible system, as we have seen, does not depend on any prior construction of a formal system of non-classical logic. . . But it does nevertheless generate a non-classical logic in a quite direct fashion:. . . So there is a sense in which the authors' approach does determine, even though it does not issue from, a non-classical logic.

why the homunculi require the stricter standard. Furthermore, the consistency criterion for components entails that *B*-structures cannot contain self-contradictory beliefs.[4] Of course, some paraconsistent systems (e.g., [Rescher and Brandom, 1979]) similarly rule out self-contradictions (generally by forbidding adjunction) but this needs some more justification, especially as the self-contradiction might not be very obvious). Finally, the requirement for local consistency will inevitably force the unnatural separation of things which are most intimately related. Just consider any proposition and its negation: they must be slotted in distinct compartments (hence the 2-partitioning requirement on inconsistent *B*-structures). But now compartments do not, in any intelligible way, capture *coherent* fragments of the theory. They most certainly do not encapsulate the error we often presume the contradiction to be. This pushes the notion of topicality out of the logical structure of the theory and into some logically arbitrary choice. While we are used to the fact that AGM-style rationality constraints are exceedingly loose (that is, there are always other considerations which must be brought into play to uniquely determine a revision operator), the uniqueness of the finest splitting is a powerful reason to treat Parikh-relevance as a significant core feature of revision. While it may be justifiable in a variety of circumstances to violate Parikh-relevance, it does capture some core intuitions about Standard Operating (revision) Procedure.

This neglect of local paraconsistency is deep-seated. For example, in their discussion of various potential relevance relations, they write:

> For instance it is not clear what intuition about relevance **R2** [that α is

[4] Presumably, we can treat *B*-structures which fail 1-consistency as a degenerate case, or as a indication of a problematically contradictory set of beliefs.

relevant to $\beta \wedge \gamma$ if and only if α is relevant to $\beta \rightarrow \gamma$] captures. If we let $\beta = p$, $\gamma = \neg p$, then $\beta \rightarrow \gamma$ 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*.

It would be hard to imagine a more direct denial of contradiction compartmentalization: no contradiction is relevant to *anything*. But surely our intuitions are pretty clear about the relevance of contradictory sentences, self-contradictory sentences, and explicitly self-contradictory sentences. For example, if I believe that all birds have feathers, no slug has feathers, and some slugs are birds, clearly these remain relevant to my beliefs about birds, slugs, and feathers regardless of whether they form a set of three sentences, one conjunction, or we add the more explicit "all birds have feathers and some birds (which are slugs) don't have feathers". Indeed, it hardly seems correct to say that $\beta \wedge \neg\beta$ is not relevant to anything, wherein both β and $\neg\beta$ are relevant to something, and remain relevant even when they are in the same set. For example, it's obvious that β and $\neg\beta$ are both very relevant to $\beta \wedge \neg\beta$.

4.2.1 Applying *B*-structures: Implicit Beliefs

B-structures are supposed to be a better model of belief states than traditional AGM belief sets. The very fact of their inconsistency tolerance is an important facet of their improved modeling of belief states. Chopra and Parikh attempt to shows this by developing a theory of *implicit beliefs*. Implicit beliefs are a special subset of the entailments of an inconsistent *B*-structure (and its corresponding theory). Chopra and Parikh explicate this via a notion of query answering. Given a query formula, A, query answering consists of two stages:

1. Find all the sublanguages (and thus subtheories) that are relevant to determining whether A. A sublanguage is relevant to a query just in case the language of the query (i.e., the smallest subset of L in which A or some formula equivalent to A may be expressed) has a nonempty intersection with that sublanguage.

2. Using Belnap's standard four values, if the union of those subtheories (Γ_A) is inconsistent, then the answer to A? is $\top$ (i.e., the query is inconsistent). If $\Gamma_A \vdash A$, then the answer to A? is yes. If $\Gamma_A \vdash \neg A$, then no, with $\bot$ otherwise.

Chopra and Parikh say that one *implicitly* believes everything that, if queried, one affirms. For an inconsistent B-structure, the affirmations will be a subset of the entailments — since, after all, the inconsistent B-structure still *entails* everything. Query answering (thus affirmation) is not directly defined in terms of the entailment of all the beliefs in the belief structure, but only of the "relevant" ones. Relevance is purely a matter of the relationship between the signature of the query and the signatures of the components of the B-structure. In this, B-structure theory is similar to splitting theory. However, the components of a B-structure — thus, the beliefs which are relevant to the query — will partially depend on logically arbitrary choices. Recall that the *only* constraint on B-structure components is that they must be consistent. They need not be minimal, nor need they encapsulate their signature. It is still the case that every component whose signature overlaps the signature of the query (or the revision input) is relevant, but this leaves a lot of room for separating contradictions out. If the contradiction in the belief base does not directly involve the signature of A, then it is often possible to find a B-structure for which A is an implicit belief.

Example 13. *Let* $\beta = \{A, B \to \neg A, C \to B, C\}$. *Let* $\beta_1 = \{\{A, B \to \neg A\}, \{C \to B, C\}\}$, $\beta_2 = \{\{A\}, \{B \to \neg A\}, \{C \to B, C\}\}$, *and* $\beta_3 = \{\{A\}, \{B \to \neg A\}, \{C \to B\}, \{C\}\}$ *(that is,* β_1*,* β_2*, and* β_3 *are distinct B-structures defined over* β*). Now consider for which of these B is an implicit belief. It is a consequence of all them, as all are inconsistent. The following table shows the relevant compartments, the state of their union (consistent or inconsistent), and the value for B (*$true, false, \top, or \bot$*) given those components:*

B-structure	*Relevant components*	*Union is?*	*Value of B*
β	$\{A, B \to \neg A, C \to B, C\}$	*inconsistent*	$\top$
β_1	$\{A, B \to \neg A\}, \{C \to B, C\}$	*inconsistent*	$\top$
β_2	$\{B \to \neg A\}, \{C \to B, C\}$	*consistent*	*true*
β_3	$\{B \to \neg A\}, \{C \to B\}$	*consistent*	$\bot$

Notice that B is an implicit belief only of β_2 *and varies from being inconsistent (in* β *and* β_2*) to being underdetermined. Essentially, in order to affirm B, we need to ensure that C gets into the selected components while A does not. The query itself, B, requires both* $B \to \neg A$ *and* $C \to B$ *no matter how they are distributed in the B-structure (since the signature of B intersects with that of both those sentences). It's easy to verify that A is true and* $\neg A$ *is* $\bot$ *in* β_1*,* β_2*, and* β_3 *(though there are B-structures over* β *for which* $\neg A$ *is true).*

Each B-structure can be seen as encoding a different relevance structure. β is the same as the finest base splitting of itself — there are explicit formulae whose signature includes B which touch explicit formulae involving every part of the overall signature. B_1 says that B is relevant to A and C separately. That is, B's relevance to each of them does not make them relevant to each other. B_2 suggests that C is more tightly relevant than A to B. And β_3 tries to keep things maximally independent. In β_3, a query formula's signature must directly contain an atom if it is to pull in *all* the formulae in β_3 with signatures containing that atom.

In example 13, β and β_3 represent two extremes: β_1 is the most "signature coherent", whereas β_3 maximizes consistency opportunities. β is closest to a split base where the thing split was itself a belief base (i.e., we were respecting the explicit syntax of the base). Of course, if we were to normalize it, we could use several explicit contradictions to capture the belief set. Perhaps something like $\{A \wedge \neg A, B \wedge \neg B, C \wedge \neg C\}$ would be appropriate (though boring and trivial, which itself may be appropriate). However, consider $\beta \setminus A$ and $\beta \setminus (B \rightarrow \neg A)$. Each of these is consistent, and we can easily find the respective normalized bases: $\{\neg A, B, C\}$ and $\{A, B, C\}$. Each of these has the same (obvious) splitting language. Interestingly, their union forms the set of implicit atomic beliefs of β_2, but *not* β_3. Not surprisingly, finer grain avoids more contradiction, but also minimizes the interesting inferential power of a B-structure. Greater refinement (as shown by Chopra and Parikh's Theorem 6.1) involves trading off some answers being implicit (because no longer inconsistent) for some answers no longer being implicit (because the components they touch no longer support them). As one might expect, there is no promise that there will be a refinement with the maximal set of implicit beliefs, nor even that there is one maximal set of implicit beliefs. If we just consider β, we see that A will be implicit in some B-structures and $\neg A$ in others, but they cannot both be implicit in the same

B-structure.

The most refined B-structure will always be the set of singleton sets each containing one formula from the base. Another way of characterizing this is by the query procedure: instead of components, a query α is evaluated against the subset of a belief base, B, such that if $b \in B$ then $\mathrm{Sig}(\alpha) \cap \mathrm{Sig}(b) \neq \emptyset$.[5] The evaluation base for any formula will, thus, be minimal. Unlike with splitting, this maximality has a strong whiff of pointlessness: if we *always* ground on individual formulae, there doesn't seem to be much *structure* left — certainly no inter-formula structure. While this leaves a role for extra-logical information to supply structure, it does so at too high a cost. Consistent B-structures may defy splitting considerations, or may not, but it is not clear why. Avoiding contradiction is a worthy goal, but *merely* avoiding contradiction seems a strange motivation for structuring a theory (without eliminating the contradictions!). Indeed, the wily airline example seems foolish as well as distasteful. In this example, the airline overbooks a 100 seat flight with 110 reservations. The airline is supposed to assert to each customer that they have a seat, but we also model a constraint that there can be only 100 customers with seats. Because we can separate all the "you have a seat" assertions and the constraint into separate components, they will all be implicit beliefs of the system. (Since each reservation supposedly only pulls in the cardinality constraint, which, if evaluated together with only one assignment, is satisfied.)

But this is clearly nonsense. Surely the relevant constraint would be that no two passengers have the same seat (except, perhaps, a baby held in a lap). Each passenger would have a seat assignment, and some seats would have two passengers. There are

[5]Notice that we are sneaking ever closer to a relevance logic *per se*.

several other ways of modeling this, but none seem nearly as odd as an airline tolerating a contradiction and saying that everything is hunky dory so long as not everyone shows up. (In point of fact, that is what happens, but then this is simply modeled as well, such that overbooking isn't contradictory. That is, we want our theory to entail that there is an overbooking, not that the theory is inconsistent because there was an overbooking.)

While we might concede, even embrace, the idea that there could be extra-logical considerations which help structure a theory, *B*-structures leave no role for the logical structure to matter. Indeed, the mere introduction of a contradiction seems to throw all splitting considerations (and thus signature coherence) to the wind.

4.2.2 Applying *B*-structures: Revision

B-structure revision involves two features of a *B*-structure: the beliefs themselves, and their organization. This generates two basic revision strategies: one which respects the antecedent structure and one which alters that structure. Chopra and Parikh refer to these as Options A and B, or "non-merging" revision and "merging" revision. In non-merging revision, the input formula revises each relevant component of the *B*-structure independently. In merging revision, first all the relevant components are gathered up, then the revision is performed on the merged component. Merged revision destroys some of the organization of the *B*-structure.

One feature of *B*-structures that we elided in our earlier discussions is the role of explicit beliefs. While a *B*-structure is a set of theories (with their associated languages), each theory (T_i) has an associated set of explicit beliefs (Γ_i). The whole *B*-

structure has an associated belief base, and, as we have seen, many B-structures can be derived from a single belief base. The role of the implicit/explicit belief distinction in [Chopra and Parikh, 1999] is not entirely clear. Clearly, given an inconsistent belief base, B, one cannot derive a sensible B-structure by first going through the corresponding belief set, $\mathrm{Cn}(B)$, as it is trivial. However, there is no need to retain the exact beliefs of the initial base. Once one has derived a set of Γ, that is, consistent sub-bases, then even though the corresponding theories T cannot be all merged together on pain of triviality, the B-structure now is independent of the Γs. For example, one could derive new characteristic formulae for the theories. Since each is consistent, there is no triviality.

At this point, a strong technical motivation for requiring consistent subtheories (or sub-bases) emerges. If each sub-base is consistent, then we can have a set of non-trivial subtheories. While we cannot, on pain of contradiction, union and take the closure of these subtheories, we can still treat each subtheory as if it were an entire belief set (at least some of the time) and thus reuse standard AGM revision operators on subtheories, instead of having to generate a new set of rationality constraints. As we saw in chapter 3, whatever you think about the intuitive basis for standard AGM theory, the rather striking inter-reduction between the various ways of conceiving revision (given by representation theorems) lends a lot of weight to them. Similarly, it is both technically and dialectically convenient to be able to drop in standard AGM operators.

In particular, this move eliminates one aspect of syntax sensitivity. Although, for some applications like ontology debugging, strict syntax *sensitivity* is critical, the general sense is that the principle of irrelevance of syntax should be adhered too insofar as is possible. B-structures have some syntax sensitivity, since they depend on logically arbitrary divisions of the base. But distinct explicit beliefs themselves can give rise to the same B-structure (if it is correct to treat the definition of B-structures as being entirely a set of language/theory pairs; if two B-structures can be significantly different solely in their explicit beliefs, then the syntax of the base clearly matters). It may be that from some bases one cannot generate certain B-structures, e.g., $\{P \wedge Q\}$ cannot generate $\{\langle\{P\}, \mathrm{Cn}(P)\rangle, \langle\{Q\}, \mathrm{Cn}(Q)\rangle\}$ without some normalization of the base; this is strictly analogous to the fact that $\{P \wedge \neg P\}$ doesn't generate a B-structure at all. In this way, B-structures exhibit a rather peculiarly *hidden* syntax sensitivity: The syntactically explicit beliefs constrain the segmentation of the signature, including whether there can be any substructure at all. Thus, there are split base driven B-structures which are not derivable from alternative bases (trivially, if you take a normalized base and they conjoin every characteristic formula, you will have a base which generates no B-structure with a component smaller than the entire theory).

These considerations lead us to tighten the definition of B-structures a bit to make clear the role of explicit beliefs:

Definition 9. *A B+-structure Bst, on a set of beliefs, B, is a set* $\{\langle S_1, B_1, T_1\rangle, \ldots, \langle S_n, B_n, T_n\rangle\}$*, such that, for all* $: i \leq n$*:*

1. $B = \bigcup B_i$;
2. $S_i = \mathrm{Sig}(B_i) = \mathrm{Sig}(T_i)$
3. $T_i = \mathrm{Cn}(B_i)$;
4. $\mathrm{Sig}(B) = \bigcup S_i = \mathrm{Sig}(\bigcup B_i) = \mathrm{Sig}(\bigcup T_i)$, *(let this be* $\mathrm{Sig}(Bst)$*);*

5. *for any B_j and B_k: $j, k \leq n$, $B_j \cap B_k = \emptyset$*

6. *for any S_j and S_k: $j, k \leq n$, it is possible that $S_j \cap S_k \neq \emptyset$*

7. *for any T_j and T_k: $j, k \leq n$, it is possible that $T_j \cap T_k \neq \emptyset$*

8. *B may be inconsistent*

   A B+-structure is strict *or* strictly consistent *iff $T_i \subset \mathrm{Cn}(\bot)$ (that is, each component theory is (individually) consistent).*

   A S_i is called a topic*, and each S_i is the* topic for *B_i and T_i.*

   A B_i is the set of explicit beliefs *about a topic S_i.*

   A T_i is the set of beliefs *about a topic S_i. If T_i is consistent, then $T_i \setminus B_i$ is the set of* implicit beliefs *about S_i.*

If we consider revision to be an operation from B+-structures to B+-structures it is clear that there are three aspects that such operations could work on corresponding to the three parts of the components of a B+-structure. For example: Let $Bst = \{\langle\{P, Q\}, \{P, \neg Q\}, \mathrm{Cn}(\{P, \neg Q\})\rangle\}$. (In general, we will leave the theory part implicit.) Now, $Bst * P \vee S$ could result in either

$$\{\langle\{P, Q, S\}, \{P, \neg Q\rangle\}$$

or

$$\{\langle\{P, Q, S\}, \{P, \neg Q, P \vee S\}\rangle\}$$

(note that, $\mathrm{Cn}(\{P, \neg Q\}) = \mathrm{Cn}(\{P, \neg Q, P \vee S\})$), if the signatures of the sets are the same and, thus, in the first set, non-minimal). These are consistent with Option A/non-merging revision. However, the revision could also result in

$\{\langle\{P, Q\}, \{P, \neg Q\}, \ldots\rangle, \langle\{P, S\}, \{P \vee S\}, \ldots\rangle\}$.

In other words, a revision operator could change the nature of a topic, and leave the corresponding explicit beliefs (though not the total beliefs) untouched. An operator could change some topic-specific set of explicit beliefs (and thus, perhaps but not necessarily, the topic). An operator could introduce a new topic with new explicit beliefs. Another possibility is that an operator might cause topics (and their related explicit beliefs) to merge, or perhaps to split. While the above example reduced to expansion, suppose instead we tried to revise Bst by $\neg P \wedge S$. In this case, it seems plausible that we should be able to achieve, with *some* operator the B+-structure $\{\langle\{P, Q\}, \{P, \neg Q\}\rangle, \langle\{P, S\}, \{\neg P \wedge S\}\rangle\}$, as well as $\{\langle\{P, Q, S\}, \{\neg P, \neg Q, S\}\rangle\}$ since we could easily derive the latter from the inconsistent base $\{P, \neg Q, \neg P \wedge S\}$. In fact, neither B-structure is a possible outcome of Chopra and Parikh's non-merging revision:

Definition 10. *[Chopra and Parikh, 1999]. For a signature, S, let the set of well-formed formulae for a given logic using S be L_S. For a formula, α, let the* shadow*,* $\mathrm{Shad}(\alpha, L_S)$*, of α on L_S (or, by abuse of notation, on S) be such that* $\mathrm{Shad}(\alpha, L_S) = \{\mathrm{Cn}(\alpha) \cap L_S\}$.

A shadow *of formula, α on a language*[6] *L_S with signature S is the intersection of the consequences of α and L_S, i.e.:* $\mathrm{Shad}(\alpha, L_S) = \{\mathrm{Cn}(\alpha) \cap L_S\}$

A characteristic formula*,* $\mathrm{Cf}(T)$*, for a theory, T, is any formula which entails the theory, i.e.,* $\mathrm{Cn}(\mathrm{Cf}(T)) = T$.

A revision operator, $*$, on a B-structure, $Bst = \{\langle S_1, T_1\rangle, \ldots, \langle S_n, T_n\rangle\}$, derived from a belief base, B, is a non-merging revision operator *just in case, for any formula α and $i \leq n$:*

[6] Where a language is all the well-formed formulae over the signature.

1. *if* $\mathrm{Cn}(\mathrm{Shad}(\alpha, L_{S_i})) = \mathrm{Cn}(\emptyset)$*, then* $\langle S_i, T_i\rangle \in (Bst * \alpha)$*,*

2. *otherwise* $\langle S_i, T_i\rangle \notin (Bst * \alpha)$ *but* $\langle S_i, T_i\rangle *_i \mathrm{Cf}(\mathrm{Shad}(\alpha, L_{S_i})) \in (Bst * \alpha)$[7].

Intuitively, a non-merging revision operator does not touch compartments with signatures that do not intersect with the signature of the input formula. If a compartment does have an relevant signature, then we revise each of those compartments with the input formula *separately* from all the others.

Example 14. *Let Bst be the B-structure*

$$\langle\{P\}, \mathrm{Cn}(\neg P)\rangle, \langle\{Q\}, \mathrm{Cn}(Q)\rangle, \langle\{Q, P\}, \mathrm{Cn}(P \vee \neg Q)\rangle\}$$

(Note that Bst is inconsistent.) Also let

$$\alpha = \neg Q$$

the result of a non-merging revision with a trivializing local revision policy is

$$\langle\{P\}, \mathrm{Cn}(\neg P)\rangle, \langle\{Q\}, \mathrm{Cn}(\neg Q)\rangle, \langle\{Q, P\}, \mathrm{Cn}(\neg Q, P \vee \neg Q)\rangle\}$$

while $Bst * \alpha$ *happens to be consistent, this was accidental. If Bst had had another compartment,* $\langle\{P\}, \mathrm{Cn}(P)\rangle$*, that would have been left untouched in any non-merging revision.*

It is clear that non-merging revision operators do not satisfy the AGM postulates as they need not satisfy any analogue of Success. They are, in fact, semi-revision operators. Chopra and Parikh, in fact, give an example that demonstrates this:

Example 15. *[Chopra and Parikh, 1999]. Let Bst be the B-structure*

$$\{\langle\{P\}, \mathrm{Cn}(P)\rangle, \langle\{Q\}, \mathrm{Cn}(Q)\rangle\}$$

and

[7] $*_i$ is any specified standard revision operator. These can vary from compartment to compartment.

$$\alpha = (P \to \neg Q)$$

With a non-merging revision operator, $(Bst * \alpha)$ *is just* Bst *again. While* α *is Parikh-relevant to both components of* Bst *(due to the non-empty intersection between* $\mathrm{Sig}(\alpha)$ *and the various* S_i*), in fact* $\mathrm{Cn}(\mathrm{Shad}(\alpha, L_P)) = \mathrm{Cn}(\emptyset)$ *(and likewise for* L_Q*). To see this, observe that* α *cannot entail any non-tautological formula with a signature smaller than* $\{P, Q\}$*. Thus, there is no component theory,* T *of* $Bst * \alpha$ *such* $\alpha \in T$*. That is,* α *is neither an explicit nor an implicit belief of* $Bst * \alpha$*. It is not even a contradictory belief in* $Bst * \alpha$*.*

Non-merging revision invariantly preserves the old topic structure. While one may be able to justify this when the signature of any input formula is a subset of the signature of the *B*-structure (that is, one might regard all inputs which are partially "off topic" as so confused that they must be rejected until reformulated in a topic appropriate way), it is very counter-intuitive in the face of signature expansion.[8] Indeed, it requires rejecting any formula involving new subject matter altogether, as well as ignoring the inherent topicality of an input.

Merging revision does attend to the topicality of the input, but still falls down on new subjects:

Definition 11. *[Chopra and Parikh, 1999]*[9]*. A revision operator,* $*$*, on a B-structure,*

[8] This is true whether one considers true signature expansion, or if one populates the *B*-structure with empty compartments for all "new" terms or, somehow, with groups of new terms. It is possible to think of signature expansion as occurring in a different stage than the revision, i.e., first one considers the new signature; then adjusts the topic structure; and then, only then, one revises. This seems a little odd as one would expect to derive insight into a new topic structure on the basis of the revision process!

[9] We note that in [Chopra and Parikh, 1999], this definition is rather strangely given in terms of explicit beliefs. This is unnecessary and not exploited at all in [Chopra and Parikh, 1999].

$Bst = \{\langle S_1, T_1\rangle, \ldots, \langle S_n, T_n\rangle\}$, *derived from a belief base, B, is a* merging revision operator *just in case, for any formulae* α *and* $i \leq n$:

- *If* $\mathrm{Sig}(\alpha) \cap S_i = \emptyset$, *then* $T_i \in Bst * \alpha$.
- *Let* $\mathcal{T}_\alpha = \{T_i | S_i \cap \mathrm{Sig}(\alpha) \neq \emptyset\}$.

  Let $T_\alpha = \mathrm{Cn}(\bigcup \mathcal{T}_\alpha)$.

  Then, $Bst * \alpha = (Bst \setminus \mathcal{T}_\alpha) \cup \{\langle \mathrm{Sig}(T_\alpha), T_\alpha * \alpha\rangle\}$.

That is, to revise *Bst* by α, first gather together all the α relevant components, union them, revise the union, then replace the relevant components with the unioned, revised one. Merging revision can change the topic organization of a *B*-structure. In principle, merging revision revises all the *beliefs* that are topic-relevant to the input as a whole — that is, as falling under a single topic induced by the input — whereas non-merging revision revises each topic separately, preserving their distinctness. Unfortunately, Chopra and Parikh do not discuss what happens when $T_\alpha = \mathrm{Cn}(\perp)$, that is, when the merge topic is inconsistent. If we follow the spirit of non-merging revision, we could simply reject such inputs (that is, fall back on semi-revision), on the grounds that we're already confused on that topic (being inconsistent) or on the grounds that no new information on that topic is possible. Alternatively, presuming that the input is not self-inconsistent, we could fall back on non-merging revision. If the topic *is* self-inconsistent, then it's not clear we can do anything but reject it, given the consistency constraint on components. Again, the inability to massage the input is a pretty severe problem.

While we might accept that these two families of operators, as with revision vs. update operators, must be selected on the basis of extra-logical evidence, it is harder to understand why new information can only alter our topic structure by *merging* topics. It is hard to see what intuitions would ground forbidding splits or mere expansions. In fact, if we take B-structures to model multi-agent scenarios (where each topic corresponds to an agent), then neither merging nor splitting make much sense (though contracting and expanding topics do), and the inability to have duplicate topics (with possibly distinct theories) likewise seems wrong. Consider a pair of contrarian recluses who live together[10]. It seems possible that a correct B-structure would have two components with the same signature (they talk to each other about everything they think about), i.e., $Bst = \{\langle S, T_1\rangle, \langle S, T_2\rangle\}$. Now suppose they are given a fresh input P such that $\neg P \notin T_1$ but $\neg P \in T_2$. Presumably, after applying non-merging revision, $P \in T_1, T_2$. But this would be incorrect. Perversity would preserve T_2 and expand T_1 by P. While an extreme case, one could also imagine a case where wherein a conflicting input was isolated in its own component of a B-structure e.g., when one is playing devil's advocate. If separating contradictory formulae was sensible in the first place (in order to generate the B-structure) it is hard to see why it would *not* sometimes be a sensible revision strategy.

Of course, definition 6 permits duplicate signatures in B-structures: they just cannot be induced by a revision. While this mis-match is disturbing, as is the priority of topic-merging over any other sort of topic modification, the mere fact of duplicate topics undercuts the whole Parikhian project. After all, in what sense do the components of a B-structure, $Bst = \{\langle S, T_1\rangle, \langle S, T_2\rangle\}$, deal with *different* topics? The *theories* may be different, but this is not, in itself, surprising. For any given topic, we know there are multiple distinct, perhaps incompatible, theories. Theories with overlapping topics also

[10]See the play *Knock Knock* by Jules Feiffer.

make sense, though generally we want a bit more structure to the overlap. For example, do two theories *agree* or *disagree* on their common topic and to what extent? Is one theory a subtheory of another? But it is not too surprising when physics overlaps with statistics and thus with economics. We might consider some notion of the *proportion* of a theory devoted to a sub-topic as a way of indicating what is "essential" vs. "incidental" to a topic. Indeed, we might consider two theories with slightly (or inessentially) differing topics to "really" be about the same (general) topic. But if the topics coincide exactly, its hard to see that the *topic* is driving the separation of the theories, especially given that merging revision is promiscuous in its merging. It doesn't try to create a new topic that covers the input sentence in some minimal way. In fact, it doesn't offer that as an option. If a component topic overlaps with the signature of the input formulae, it is pulled in.

While in non-merging revision, topicality does play a strong role (indeed, forcing semi-revision), the overall picture is that theories, not signatures, are the primary object of consideration. After all, the prime considerations for division are consistency (of the theory), which is a strict condition, and the set of implicit beliefs. Topicality itself plays an incidental role in the generation of sub-theories.

Returning to B+-structures, we can see several advantages. Since we maintain base formulae, we can accept self-inconsistent inputs. Since we allow topics to be altered, we can merge, split, or add topics as seems appropriate, which allows us to accept inputs with new signatures, or inputs that span more than one topic (either by merging, or creating a new topic). In general, if we shift our attention from topicality to *segmentation* of belief bases, we easily arrive at Sven Ove Hansson and Renata Wassermann's theory of *local change*.

4.3 Local Change

In [Hansson and Wassermann, 2002], we find an account of local theory change operations (including, revision, semi-revision, contraction, and consolidation) in terms of a localized consequence relation. This relation, and the associated change operations, is defined strictly in terms of belief bases and their associated (global) consequence relation.[11]

As in splitting and B-structures theory, Hansson and Wassermann wish to divide belief bases into smaller components. As in splitting theory (though less so with B-structures) they wish these divisions to be driven by the logical structure of the belief base:

> Compartments of belief bases can be seen as representations of compartments of our minds or databases. There are two major ways to introduce them. First, they may be introduced as an addition to the logic, so that one and the same belief base can be divided into compartments in different ways. Secondly, they may be derived from the logic. The second method is the more economical, requiring no extra entities, and should be tried out first. We are going to use it here.

B-structures introduce new features to the logic, to wit, the particular split of the language. Clearly, the very same belief base can generate many different B-structures, and

[11] They briefly describe the motivation for belief bases: essentially, better computability and greater expressive power — in particular, the ability to distinguish between inconsistent belief states. The only downside they mention is increased complexity over belief sets ([Hansson and Wassermann, 2002], pg. 50.).

not merely in a refinement hierarchy. Different (inconsistent) B-structures have different consequence relations (i.e., sanction different implicit beliefs) and react differently to revision.

While logical structure does drive membership in a compartment, unlike with splitting theory, logical structure *alone* does not drive membership. As with B-structures, some choices are made based on logically arbitrary considerations. Unlike B-structures, after the initial choice of a "seed", the rest of a compartment is determined by logical structure alone. The seed of a compartment is a subset of the belief base. Compartments for sets with cardinality greater than one can be defined as the union of the compartments of each formula. Intuitively, the compartment of a formula in a belief base is the set of all the relevant formulae in the belief base. Here, these are the sentences which entail the compartmentalized formula or its negation, with the proviso that contradictions and tautologies are not relevant to anything:

Definition 12. *[Hansson and Wassermann, 2002] (definition 2.3, quoted exactly) Let C be an inference operation. The function c is the* compartmentalization function *based on C if and only if, for all $A, B \subseteq L$[12]: $c(A, B) = \bigcup_{\alpha \in A} c(\alpha, B)$, where*

$c(\alpha, B) = \emptyset$, *if* $\alpha \in C(\emptyset)$ *or* $\neg\alpha \in C(\emptyset)$

$c(\alpha, B) = \bigcup((B \perp\!\!\!\perp_C {}^{13}\alpha) \cup (B \perp\!\!\!\perp_C \neg\alpha) \setminus (B \perp\!\!\!\perp_C \bot)$, *otherwise.*

In other words, like B-structures, compartments cannot contain self-inconsistent

[12] Presumably, L is the set of all formulae, i.e., the language.

[13] $\perp\!\!\!\perp$ is the *kernel* operator, parameterized to a consequence relation, C. Essentially, $B \perp\!\!\!\perp_C \alpha$ is the set of minimal subsets of B which entail α according to C. Kernels are also known as *justifications* [Kalyanpur et al., 2007].

formulae, even relevant ones. For example, the compartment of P in $B1 = \{P \wedge \neg P, \neg P \wedge (Q \wedge \neg Q), R \wedge \neg R\}$ is the empty set. Yet, the compartment for P in $B2 = \{P, \neg P, \neg P \wedge Q, \neg Q, R \wedge \neg R\}$ is $\{P, \neg P, \neg P \wedge Q\}$. In the first case, we seem to be missing crucially relevant formulae, although we do exclude the clearly irrelevant $R \wedge \neg R$. In the second case, one might be disturbed by the fact that the compartment is not complete with regard to its signature. In splitting theory, having Q in the signature is enough to pull in all the formulae with Q in their signatures. Clearly, there will be overlap between $c(P, B2)$ and $c(Q, B2) = \{\neg P \wedge Q, \neg Q\}$, but, unlike with splitting theory, there is no way to determine topicality from mere inspection of the component or its signature. Instead, we need to inspect the original base to determine, given the base, what, e.g., $c(Q, B2)$ *could* be a compartment for.

Hansson and Wassermann follow the standard pattern of defining local consistency and local implication in terms of compartments, and then defining the (base) revision operators on top of these. One interesting feature of [Hansson and Wassermann, 2002] is that the definition of local change operations is nearly an afterthought. Instead of defining local operations directly, they first give very general definitions of the various change operations (in their case, contraction, consolidation, external and internal revision, and semi-revision). The definitions are general in two respects: First, they abstract away, where possible, the distinction between belief base vs. belief set change.[14] Second, they

[14]In some cases, a change operation doesn't really make sense in the belief set, aka, theory context. Consolidation and thus external revision are the obvious cases. Consolidation isn't meaningful for standard belief sets (i.e., those defined by an explosive consequence relation) because all contradictory belief sets are identical, reducing the "choice" of how to contract by a contradiction to mere selection of an arbitrary

rely on very general properties of a consequence relation. For example, in definition 3.2 of [Hansson and Wassermann, 2002], the kernel contraction operator is defined in terms of an arbitrary inference operation CR, then theorem 3.3 (their generalized representation theorem for their contraction operation) begins, "Let C be an inference operation satisfying monotony and compactness. Then $\div$ is an operation of kernel contraction on B determined by C and some incision function if and only if for all sentences α...[a series of AGM style axioms]." As long as localized implication satisfies monotony and compactness, it defines a set of (kernel) contraction operations.

Unfortunately, the local change operators do not apply to B-structures in the general case, since the components of a B-structure (or B+-structure) do not have to be compartments. This is easily seen:

Example 16. *In the B-structure:*

$$\{\langle\{P\}, \mathrm{Cn}(P)\rangle, \langle\{P, Q\}, \mathrm{Cn}(Q \to \neg P, Q)\rangle\}$$

neither component is a compartment for P, for ¬P, for Q, or for $(Q \to \neg P)$ *(if we take the underlying base to be* $\{P, Q \to \neg P, Q\}$*). For the first two, both components are too small (i.e., they do not contain the kernels for the complementary literal). For the second two formulae, the second component is too big, i.e., contains things which could be removed without breaking the entailment.*

B-structures, being based on (logically) arbitrary decisions, can easily come apart from compartments. Furthermore, compartments are *relative* to a seed. What's more or less stable is the compartmentalization function — the compartments returned from

consistent set of sentences from the language. External revision relies on consolidation.

that function depend on the second argument. So any static division, even if it generates compartments for some cases, invariably will fail to capture compartments for other cases. Thus, local change has three points of arbitrariness:

1. It depends on the syntactic form of the base. Equivalent bases can give radically different revision results.

2. Compartments depend on the seed in question.

3. Seeds in question depend on the sentences to be retracted.

Local change is always merging — compartments pull in *everything* relevant to the input as a unitary whole. Local change is also, surprisingly, computationally reasonable. Intuitively, computing kernels is no harder than reasoning itself (assuming that reasoning is NP-hard), and, for many logics, we have good empirical evidence that computing kernels (i.e., justifications) is reasonable (see, [Kalyanpur et al., 2007]). (How difficult the overall procedure is depends on the change operators). This reasonableness claim is a little different that Parikh's computational claims: We observe that as long as we *can reason* with a theory, we can (in all probability) compute the kernels as well. That is, we accept the worst case complexity of the logic, but observe that we can, for many real cases, do just fine with it. That is, compartmentalization does not *alleviate* the complexity, but it also doesn't make it *worse*.

However, the fact that we cannot determine a compartment without knowing the input is, prima facie, quite damaging to the ability of local change to underwrite an account of topicality and topic oriented revision. To put it bluntly, local change does not provide any sort of analysis of the object of revision. With splitting theory, we had an intuitively plausible criterion of topicality (disjointness of signature) and some reinforcing results (univocal finest splittings). With *B*-structures, topicality was extra-logically determined, but clearly *there*. Indeed, the structure of *B*-structures given by topics drove different fundamental sorts of revision operators. Local change says nothing *until* you try to revise and then we don't have topicality so much as exact relevance. In spirit, compartments are very similar to maximal non-entailing subsets, minimal entailing subsets, and all the other sorts of apparatus AGM style theories have mobilized. (This shouldn't be too surprising given that kernels are a basic part of base revision.) Thus, we aren't really exploiting the structure of the theory in any interesting way.

4.4 Parikhian Compartments

While disappointing as a theory of topical revision, local change does provide some apparatus that allows us to rehabilitate splitting theory to meet the critique presented in [Chopra and Parikh, 1999] without the failings of *B*-structures(or even *B*+-structures). Let us consider consistent theories first. It is clear that the finest splitting of a theory (or a base) is a collection of compartments for the set of all[15] topic-confined inputs. Recall that

[15] A split can contain components that are not minimal with respect to specific inputs confined to the corresponding language, and typically will. But this is exactly what one would expect when considering the *topic* of an input, rather than the specific content of an input. For example, if my beliefs entail that I believe that my refrigerator is sentient, it's entirely possible that the kernals supporting that belief (or its negation) omit beliefs that support my implicit belief that my refrigerator has a door. But if I am considering everything that I believe about the *topic* of my refrigerator, I would to expect to consider both its sentience and its door.

each component theory (or base) of a splitting is sufficient to capture *all* the entailments of the parent theory which are expressed in terms of that splitting's signature. That is, a splitting element captures all the entirely on-topic implicit beliefs of a theory. This follows directly from parallel interpolation. Thus, for any (consistent) theory (or base), B, and any contraction input with a signature confined to some element B_i, α, the kernels of α in B are all going to be subsets of B_i. So they aren't *quite* compartments, since they are somewhat larger. But this is what we'd expect from a topic account! Not every on-topic entailment follows from the *whole* topic — i.e., topics have substructure. This distinguishes topicality from relevance. Not every on topic sentence is relevant to arbitrary input.

Similarly, if we have a cross-topic formula to contract, we need merely take all the topics whose signature intersects with the signature of the input to form a (superset of a) compartment for that input. This is, of course, exactly what we do for Parikhian revision under Axiom P. Whether we choose to make our topic focused revision operator a local change operator as well depends on the situation. If we are modeling situations where the topic is well accepted and considered to be stable, reliable, or well entrenched, we might well choose to localize the change as much as possible. Similarly, if the overall theory is shaky, new, or otherwise unstable, we might prefer to let new inputs have more radical and far reaching (yet still topic oriented) effects.

If we consider inconsistent theories, we can still point to univocal splittings if we simply consider split (unnormalized) bases. That is, we divide the base into a series of minimal subsets with disjoint signatures. At this point, we will (potentially) have both consistent and inconsistent components. For the consistent ones, the above arguments hold (i.e., they are appropriate supersets of the union of compartments for all entailments expressible in the signature of that component, i.e., "on-topic" entailments). For inconsistent components, presuming no self-inconsistent formulae, it is easy to see that they are supersets of the relevant compartments as well. Given that there are no self-inconsistent statements, we can consider only consistent sentences α such that $(B \perp\!\!\!\perp \alpha)$ $(B \perp\!\!\!\perp \bot)$ is non-empty. That is, we focus only on formulae which do not follow from the inconsistency of B alone. Now, suppose that B_i is the only inconsistent component of B. We know that $\bigcup(B \perp\!\!\!\perp \alpha) \subseteq B_i$ because, by splitting, $\mathrm{Sig}(\alpha) \subseteq \mathrm{Sig}(B_i)$. By parallel interpolation, no formula with a disjoint signature can entail (or be part of a minimal support for) α. The same principle applies for $\neg\alpha$. Thus, B_i is a superset of the compartment for α and the principles of local change apply.

We now have a theory of topicality (minimal disjoint subsets of the base) for which topics are univocal, entirely logic-driven, and reasonably computable. Furthermore, we have a bog-standard account of revision (local change) which, with slight modifications, applies to our topic structured bases. Our theory of topicality confines contradictory beliefs, though, like other accounts, it does not handle self-contradictions.

Chapter 5

Conclusion

In this thesis, we have investigated a series of accounts of logically determined topicality and their relation to revision operators. One interesting fact we discovered is that topic-respecting operators are, in general, sanctioned by broader sorts of AGM (or base) postulates. Parikh's idea that topic based considerations give a strong basis for revising one's account of the rationality of operators seems seriously misguided. Aside from the fact that the various sorts of topicality we investigated simply had less radical implications than the authors sometimes claimed, it should be clear that sometimes topic-relevance isn't an overridingly rational consideration. Topicality is *one* consideration that comes into play in various circumstances and it is not irrational to consider topic organization when one is revising one's beliefs or editing a book or modifying a database. However, there are considerations which cut across topicality, for example, the origin of a belief. Similarly, not all revisable structures benefit from strict topical organization: Consider an introductory book which covers lots of topics in a pedagogic, or merely aesthetically, effective way. Or just consider a wide-ranging conversation that touches on many topics via digressions. Such things are clearly not second-order irrational (i.e., there are good reasons — pedagogy, aesthetics, comfortableness — for not being strictly topical) and, arguably, they shouldn't be first order irrational (i.e., *analogical* reasoning crosses topics; provenance crosses topics; etc.).

We have developed a logic-based notion of topicality that:

- is intuitive (i.e., disjoint signatures indicate disjoint topics);
- works for inconsistent belief bases (i.e., topics capture and confine their errors);
- is univocally determined by logically considerations alone;
- is computable;
- selects reasonable revision operators (by borrowing from the theory of local change; this also makes it applicable to a very wide range of logics).

As such, it meets more and more essential desiderata for a theory of topicality than splitting theory, *B*-structures, or local change. However, it still fails on some fronts:

- As with both local change and *B*-structures, it does not handle self-inconsistency at all. While one might argue (as many do) that self-inconsistency is inherently irrational, we have argued strenuously against that view. At the very least, a theory of topicality should be able to identify the topic of a self-inconsistent statement!
- Unlike splitting theory, our notion, being base dependent, is strongly syntax sensitive (which also drives the prior issue).
- Our topics do not overlap, nor do they say much, if anything, about topic substructure. Coupled with syntax sensitivity, we get topicality being less minimal than we would prefer. For example, with consistent theories, normalized bases give us smaller, more focused topics. On the other hand, this syntactic sensitivity allows us

to couple otherwise separate signatures without having to add extra syntax to our language.

While coping with all these issues is beyond the scope of this thesis, we can identify several promising approaches to solving them.

Going in reverse order, the key point to notice about splittings is not their disjointness, but the effect disjointness has upon entailments. Originally, of course, Parikh was motivated by the intuition that, at least, disjointness of signature was a sufficient condition for distinct topics (that is, there is no common topic between partitions with disjoint signatures). That still seems correct. But the question is *why*? If we consider parallel interpolation, the answer seems to be that a partition captures all entailments expressed in the language of the partition alone. Furthermore, any union of partitions captures all entailments expressed in the language of that union. Intuitively, if we have a disjoint partition that contains sentences about hedgehogs, we know that the rest of the belief set or base has nothing else to say about hedgehogs *alone*. If we have another partition that contains sentences about the Swedish heavy metal rock band, Opeth, we know that anything our belief set (or base) has to say about hedgehogs and Opeth alone is captured by the union of their respective partitions.

While signature disjointness is sufficient (given the appropriate logic) for separable topics it is not clear that it is necessary. The idea of "capturing all entailments" is naturally parameterized by the consequence relation underlying those entailments. We can vary the consequence relation in a number of ways:

- We can vary the signature of entailments.

- We can vary the syntactic structure of entailments. For example, for databases we might only care about entailing ground answers to conjunctive queries, or ground answers to unions of conjunctive queries, or even arbitrary first order answers to unions of conjunctive queries. Different sorts of query may be more or less sensitive to the structure of the models of the belief set (or base).

- We can vary the entailment relation by changing the semantics of our logic or adding new, unsound inference rules.

The last possibility is already accounted for by the rather broad conditions we have imposed on our logics i.e., restricted equality, monotony, supra-classicality, etc. Furthermore, if we change the *logic* of our beliefs we should expect a change in the *logical* structure, and thus the logical topicality of our beliefs.

The second is similarly uninteresting and can be seen as a variant of the first with the exception that more sensitive queries may affect the topic analysis.

The first is critical. We have already seen how queries can force otherwise distinct topics to come together. Similarly, we've also seen how narrow focus on the query (i.e., for compartments) can give us finer grain structures. The question is whether there is a natural way of distinguishing the signature for a *class* of queries in a principled way such that the queries would induce a different, perhaps overlapping set of topics. In combination with restrictions on the syntactic structure of queries, this might be quite effective. For example, if we restrict ourselves to atomic class conditional queries (i.e., in description logic lingo, atomic subsumptions between concepts, aka, unary predicates) then binary predicates (aka, roles) will never appear in any query. They still might affect topicality since they can affect atomic subsumptions. But since they don't appear in any possible query, it is at least possible that they can be shared between otherwise disjoint topics.

Syntax sensitivity was introduced in order to cope with inconsistency. We need to be able to distinguish between objects of revision that are, themselves, inconsistent. However, radical syntax sensitivity isn't necessary, even without switching to a full-blown paraconsistent consequence relation. For example, we could stick with belief bases but require that they be normalized. For example, following [Horridge et al., 2008], we could apply the well known structural transformation [Plaisted and Greenbaum, 1986] to our belief base. The structural transformation separates all "parts" of a formula, including all nested substructure. It's a common alternative to conjunctive or disjunctive normal form. One consequence of this separation is that each complementary literal appears in their own characeristic formaula. Thus, after transformation, there are no self-contradictions. There are also no syntactically arbitrary couplings (e.g., random conjunctions of otherwise unrelated literals). We could extend this further by allowing certain classes of entailment of individual sentences (e.g., equivalents). It seems reasonable that we could achieve a considerable degree of syntactic freedom in this way.

Appendix A

A preliminary experiment

A.1 Empirical Issues

Even though there are strong theoretical considerations in favor of disjointness of signature between two subsets of a theory (in appropriate logics) as a sufficient condition (at least) for logical topic separateness, if real theories tend to be highly connected then there are two obvious families of conclusion to draw: 1) that the analysis is correct but logical topicality isn't a comment feature of such theories or 2) that the analysis is wrong and we should look elsewhere.

To this end, a simple experiment would be to take a corpus of formalized theories and examine at least the first approximation of logical topicality. This appendix describes the results of such an experiment.

A.2 Description

The TONES ontology repository at the University of Manchester[1] is a collection of ontologies (in the computer science sense, aka, logical theories in a restricted syntax) published on the open Web by a variety of parties including NASA (about earth science) and the National Cancer Institute. The corpus is not randomly gathered, but was assembled by the maintainers from "likely" sources of "interesting" (esp. from a tool and logical ser-

[1] `http://owl.cs.manchester.ac.uk/repository/browser`

vice point of view) ontologies. The repository contains (as of this writing) 230 ontology with a wide range of size, complexity, and minimal background logic.

For this experiment, we implemented a naive version of base splitting without any normalization and applied it to the ontologies from the repository. Even so, the time to split an ontology, esp. larger ones, was dominated by download time.

The goal is to see if naturally occurring ontologies exhibit significant disjoint topic structure.

A.3 Results

Of the 230 ontologies in the repository, we were able to download and process 199. Of these, 132 had only one partition, indicating no significant disjoint substructure. Of the remaining 67, 58 had between 2 and 9 disjoint partitions (weighted toward the low end), and the last nine had up to 25 disjoint partitions (see figure A.1)

A.4 Analysis

While the large majority of ontologies showed no disjoint subsigntures, a reasonable number did which does show that Parikhian topicality is not an unknown phenomenon. Several factors suggest a bias against disjoint topicality being present in the sample and detectable by our methodology:

1. We used a rather crude approximation method: Splitting the explicit base. We made no attempt to normalize the base so it is possible that many of the ontologies had a finer grain substructure.

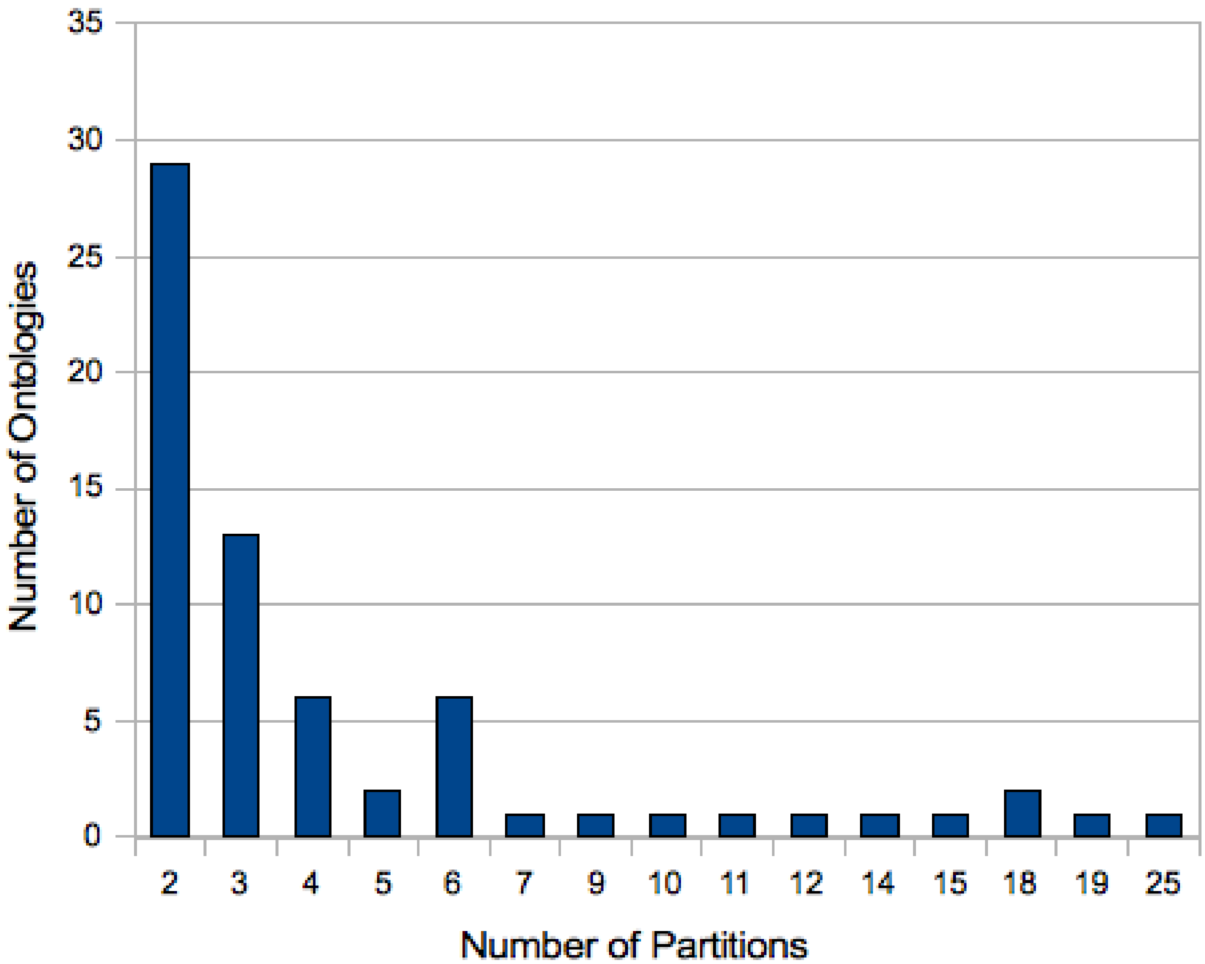

Figure A.1: This graph does not include the 132 ontologies with no topic substructure.

2. Ontologies themselves are often "single topic'. Most of the ontologies in the repository have disjoint signatures and, thus, themselves could be seen as exhibiting topic structure with respect to each other. Of course, syntactic considerations alone incline this way (i.e., uncoordinated parties don't stumble across the same spellings), so this would require some analysis of the content of terms and perhaps mapping attempts between ontologies with related or overlapping domains.

3. In some ontologies (Galen is a good example) the developers used user vocabulary to simulate functions that should be part of the logical or linguistic apparatus. For example, Galen introduced a "top class" part of whose function was to provide a top node in a browser. These sorts of modeling hack can destroy logically detectable topicality.

4. It's also possible that the particular linguistic affordances of the formalisms used encourage a non-topical style.

Overall, we believe this experiment provides limited, prima facie support for the general analysis in this thesis. As future work, we intend to delve further into the apparently non-topic structured ontologies to see whether they are intuitively cohesive, to develop algorithms which compute the normalized base, and to further investigate the possibility of subtopics and overlapping topics.

We should point out that one area that experiment is unhelpful is in providing insight into what happens when there are contradictions. Published OWL ontologies tend to be contradiction free for several reasons including that contradictions are repaired before publishing and OWL makes it fairly easy to produce nominally consistent ontologies.

Bibliography

[Alchourrón et al., 1985] Alchourrón, C., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, (50):510–530.

[Alchourrón and Makinson, 1985] Alchourrón, C. and Makinson, D. (1985). On the logic of theory change: Safe contraction. *Studia Logica*, (44):405–422.

[Allen, 1974] Allen, F. E. (1974). Interprocedural data flow analysis. In *IFIP Congress*, pages 398–402.

[Amir and McIlraith, 2005] Amir, E. and McIlraith, S. A. (2005). Partition-based logical reasoning for first-order and propositional theories. *Artif. Intell.*, 162(1-2):49–88.

[Baader, 1999] Baader, F. (1999). Logic-based knowledge representation. In Wooldridge, M. J. and Veloso, M., editors, *Artificial Intelligence Today, Recent Trends and Developments*, pages 13–41. Springer Verlag.

[Baader and Hollunder, 1995] Baader, F. and Hollunder, B. (1995). Embedding defaults into terminological knowledge representation formalisms. *J. Autom. Reasoning*, 14(1):149–180.

[Bienvenu et al., 2008] Bienvenu, M., Herzig, A., and Qi, G. (2008). Prime implicate-based belief revision operators. In *18th European Conference on Artificial Intelligence (ECAI'08), poster*.

[Bochman, 1998] Bochman, A. (1998). A foundationalist view of the AGM theory of belief change. *Artificial Intelligence*, 116:2000.

[Bochman, 1999] Bochman, A. (1999). Entrenchment versus dependence: Coherence and foundations in belief change. *Journal of Logic, Language and Computation*.

[Brachman, 1983] Brachman, R. (1983). What is-a is and isn't: An analysis of taxonomic links in semantic networks. *IEEE Computer*, 16:30–36.

[Brewka and Augustin, 1987] Brewka, G. and Augustin, D. S. (1987). The logic of inheritance in frame systems. In *In Intl. Joint Conference on Artificial Intelligence*, pages 483–488. Morgan Kaufmann Publishers.

[Buchanan and Shortliffe, 1985] Buchanan, B. G. and Shortliffe, E. H., editors (1985). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA.

[Chopra and Parikh, 1999] Chopra, S. and Parikh, R. (1999). An inconsistency tolerant model for belief representation and belief revision. In Dean, T., editor, *IJCAI*, pages 192–199. Morgan Kaufmann.

[Collins and Quillian, 1969] Collins, A. M. and Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:240–247.

[Davis et al., 1993] Davis, R., Shrobe, H. E., and Szolovits, P. (1993). What is a knowledge representation? *AI Magazine*, 14(1):17–33.

[Doyle, 1992] Doyle, J. (1992). Reason maintenance and belief revision: Foundations vs. coherence theories. In *Belief Revision*, pages 29–51. Cambridge University Press.

[Fodor, 1983] Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.

[Gärdenfors, 1988] Gärdenfors, P. (1988). *Knowledge in Flux : Modeling the Dynamics of Epistemic States*. MIT Press.

[Gärdenfors, 1990] Gärdenfors, P. (1990). The dynamics of belief systems: Foundations vs. coherence theories. *Revue Internationale de Philosophie*, 44:24–46.

[Gärdenfors, 2000] Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. The MIT Press.

[Garson, 1989] Garson, J. (1989). Modularity and relevant logic. *Notre Dame Journal of Formal Logic*, 30(2):207–223.

[Hammer and Kogan, 1996] Hammer, P. L. and Kogan, A. (1996). Essential and redundant rules in horn knowledge bases. *Decis. Support Syst.*, 16(2):119–130.

[Hansson, 1999a] Hansson, S. O. (1999a). A survey of non-prioritized belief revision. *Erkenntnis*, 50:413–427.

[Hansson, 1999b] Hansson, S. O. (1999b). *A Textbook of Belief Dynamics : Theory Change and Database Updating (Applied Logic Series)*. Springer.

[Hansson and Wassermann, 2002] Hansson, S. O. and Wassermann, R. (2002). Local change. *Studia Logica*, 70(1):49–76.

[Hayes, 1979] Hayes, P. J. (1979). The logic of frames. In Metzing, D., editor, *Frame Conceptions and Text Understanding*, pages 46–61. Walter de Gruyter and Co., Berlin, Germany.

[Horridge et al., 2008] Horridge, M., Parsia, B., and Sattler, U. (2008). Laconic and precise justifications in OWL. In *Proc. of ISWC*.

[Horrocks et al., 2005] Horrocks, I., Parsia, B., Patel-Schneider, P. F., and Hendler, J. A. (2005). Semantic web architecture: Stack or two towers? In Fages, F. and Soliman, S., editors, *PPSWR*, volume 3703 of *Lecture Notes in Computer Science*, pages 37–41. Springer.

[Horty, 1994] Horty, J. F. (1994). Some direct theories of nonmonotonoc inheritance. In Gabbay, D. M., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic in Artificial Intelligence and Logic Programming — Nonmonotonic Reasoning and Uncertain Reasoning(Volume 3)*, pages 111–187. Clarendon Press, Oxford.

[Jaśkowski, 1969] Jaśkowski, S. (1969). Propositional calculus for contradictory deductive systems. *Studia Logica*, 24:143–160.

[Kalyanpur et al., 2007] Kalyanpur, A., Parsia, B., Horridge, M., and Sirin, E. (2007). Finding all justifications of OWL DL entailments. In Aberer, K., Choi, K.-S., Noy, N. F., Allemang, D., Lee, K.-I., Nixon, L. J. B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., and Cudré-Mauroux, P., editors, *ISWC/ASWC*, volume 4825 of *Lecture Notes in Computer Science*, pages 267–280. Springer.

[Kashyap and Borgida, 2003] Kashyap, V. and Borgida, A. (2003). Representing the UMLS semantic network using OWL: (or "What's in a Semantic Web link?"). In Fensel, D., Sycara, K. P., and Mylopoulos, J., editors, *International Semantic Web Conference*, volume 2870 of *Lecture Notes in Computer Science*, pages 1–16. Springer.

[Katsuno and Mendelzon, 1991] Katsuno, H. and Mendelzon, A. (1991). On the difference between updating a knowledge base and revising it. In Allen, J. F., Fikes, R., and Sandewall, E., editors, *KR'91: Principles of Knowledge Representation and Reasoning*, pages 387–394. Morgan Kaufmann, San Mateo, California.

[Kifer et al., 1995] Kifer, M., Lausen, G., and Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42:741–843.

[Kifer and Subrahmanian, 1992] Kifer, M. and Subrahmanian, V. S. (1992). Theory of generalized annotated logic programming and its applications. *J. Log. Program.*, 12(4):335–367.

[Kourousias and Makinson, 2006] Kourousias, G. and Makinson, D. (2006). Parallel interpolation, splitting, and relevance in belief change. *Association for Symbolic Logic*.

[Levesque and Brachman, 1987] Levesque, H. J. and Brachman, R. J. (1987). Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93.

[Liberatore, 2000] Liberatore, P. (2000). Compilability and compact representations of revision of horn knowledge bases. *ACM Trans. Comput. Logic*, 1(1):131–161.

[Liberatore, 2001] Liberatore, P. (2001). Monotonic reductions, representative equivalence, and compilation of intractable problems. *J. ACM*, 48(6):1091–1125.

[Mackie, 1977] Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin Books, New York.

[Makinson, 1982] Makinson, D. (1982). Review: Nicholas Rescher, Robert Brandom, The Logic of Inconsistency. A Study in Non-Standard Possible-World Semantics and Ontology. *Journal of Symbolic Logic*, 47(1):233–236.

[McCarthy, 1958] McCarthy, J. (1958). Programs with Common Sense. In *Proceedings of the Teddington Conference on the Mechanisation of Thought Processes*, pages 77–84.

[McDermott, 1976] McDermott, D. (1976). Artificial intelligence meets natural stupidity. *SIGART Bull.*, (57):4–9.

[McDermott, 1987] McDermott, D. (1987). A critique of pure reason. *Computational Intelligence*, 3(3):151–237.

[Minsky, 1974] Minsky, M. (1974). A framework for representing knowledge. Cambridge, MA, USA.

[Parikh, 1999] Parikh, R. (1999). Beliefs, belief revision, and splitting languages. In *Logic, language and computation, vol. 2*, pages 266–278, Stanford, CA, USA. Center for the Study of Language and Information.

[Plaisted and Greenbaum, 1986] Plaisted, D. A. and Greenbaum, S. (1986). A structure-preserving clause form translation. *Journal of Symbolic Computation*.

[Reiter and de Kleer, 1987] Reiter, R. and de Kleer, J. (1987). Foundations of assumption-based truth maintenance systems: Preliminary report. In *AAAI*, pages 183–189.

[Renz et al., 2000] Renz, J., Rauh, R., and Knauff, M. (2000). Towards cognitive adequacy of topological spatial relations. In Freksa, C., Brauer, W., Habel, C., and Wender, K. F., editors, *Spatial Cognition*, volume 1849 of *Lecture Notes in Computer Science*, pages 184–197. Springer.

[Rescher and Brandom, 1979] Rescher, N. and Brandom, R. (1979). *The Logic of Inconsistency: a Study in Non-Standard Possible World Semantics and Ontology*. American Philosophical Quarterly.

[Rott, 1992] Rott, H. (1992). On the logic of theory change: More maps between different kinds of contraction function. In Gärdenfors, P., editor, *Belief Revision*, volume 29, pages 52–88. Cambridge University Press, Cambridge, UK.

[Sioutos et al., 2007] Sioutos, N., de Coronado, S., Haber, M. W., Hartel, F. W., Shaiu, W.-L., and Wright, L. W. (2007). NCI thesaurus: A semantic model integrating cancer-related clinical and molecular information. *Journal of Biomedical Informatics*, 40(1):30 – 43. Bio*Medical Informatics.

[Strube, 1992] Strube, G. (1992). The role of cognitive science in knowledge engineering. In *Proceedings of the First Joint Workshop on Contemporary Knowledge Engineering and Cognition*, pages 161–174, London, UK. Springer-Verlag.

[Thomason and Horty, 1989] Thomason, R. H. and Horty, J. F. (1989). Logics for inheritance theory. In *Proceedings of the 2nd international workshop on Non-monotonic reasoning*, pages 220–237, New York, NY, USA. Springer-Verlag New York, Inc.

[Woods, 1975] Woods, W. A. (1975). What's in a link: Foundations for semantic networks. In Bobrow, D. G. and Collins, A. M., editors, *Representation and Understanding: Studies in Cognitive Science*, pages 35–82. Academic Press, New York.