A

Belief Structures and Sequences: Relevance Sensitive, Inconsistency Tolerant Models for Belief Revision

by

Samir Chopra

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2000

UMI Number: 9959172

UMI

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

APPROVAL

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jan 26, 2000
Date

R. Parikh
Chair of Examining Committee

1/27/2000
Date

Michael Devitt
Executive Officer

Examining Committee:

Professor Rohit Parikh
Professor Melvin Fitting
Professor Henry Kyburg Jr.
Professor Jonathan Adler
Professor Horacio Arlo-Costa

The City University of New York

ABSTRACT

Belief Structures and Sequences: Relevance Sensitive, Inconsistency Tolerant Models for Belief Revision

By

Samir Chopra

Advisor: Professor Rohit Parikh

This thesis proposes and presents two new models for belief representation and belief revision. The first model is the B-structures model which relies on a notion of partial language splitting (based on [80]) and tolerates some amount of inconsistency while retaining classical logic. The model preserves an agent's ability to answer queries in a coherent way using Belnap's four-valued logic. Axioms analogous to the AGM axioms hold for this new model. The distinction between implicit and explicit beliefs is represented and psychologically plausible, computationally tractable procedures for query answering and belief base revision are obtained.

The second model presents a method for relevance sensitive non-monotonic inference from belief sequences which incorporates insights pertaining to prioritized inference and relevance sensitive, inconsistency tolerant belief revision. Our model uses a finite, logically open sequence of propositional formulas as a representation for beliefs and defines a notion of inference from maxiconsistent subsets of formulas guided by two orderings: a temporal sequencing and an ordering based on relevance relations between the conclusion and formulas in the sequence. The relevance relations are ternary (using context as a parameter) as opposed to standard binary axiomatizations. The inference operation thus defined easily handles iterated revision by maintaining a revision history, blocks the derivation of inconsistent answers

from a possibly inconsistent sequence and maintains the distinction between explicit and implicit beliefs. In doing so, it provides a finitely representable formalism and a plausible model of reasoning for automated agents.

Dedicated to my parents.

Pramod and Prabha.

who. I think, would have been proud.

ACKNOWLEDGEMENTS

I would like to acknowledge my deep personal and intellectual debt to Professor Rohit Parikh for his help with my thesis and doctoral study. Any formal elegance noticed in the pages to follow is entirely due to him. I thank him, too, for all his financial support, the endless cups of coffee, all the lunches and most of all, patience. I thank Professors Henry Kyburg Jr., Jonathan Adler, Melvin Fitting and Horacio Arlo-Costa for their detailed and careful readings of a preliminary draft of the thesis. I would like to thank Konstantinos Georgatos for all his work on the sequences model and Graham Priest for helping to get me started with the references he suggested.

I would like to thank my friends: Brad Armour-Garb for all his help with the written examinations; Jared Blank for his generous thesis defense gift; Tom Connell for all the ballgames we watched together at Yankee Stadium and for helping to organize my life with his shelving skills; David Coady for talking to me about cricket, politics and the movies all these years; Jennifer Fisher for helping me celebrate my defense; Leanne Habeeb for reading my papers, quizzing me for the orals and for all her help with too many things to list here; Martin Harvey for lending me copies of his preparation materials; Russell Marcus for being an inspiring logician; Michael Menser for being someone I could go headbanging with; Rakesh Kushwaha, Gurinder Johar, Amit Sadana for being fellow rowdies all these years and Sumit Sadana for fixing my PC at a crucial time in my writing.

Material from this thesis was presented at talks at the Annual Meeting for the Association of Symbolic Logic, the Boston Logic Conference, the International Joint Conference on Artificial Intelligence, the Pan-Hellenic Logic Symposium, the Logic and Artificial Intelligence Seminar at the University of Maryland, the Seminar on Computational Game Theory at the Courant Institute, the New York City Logic Conference and the IBM TJ Watson Research Center. I thank the audiences at these fora for their helpful comments. In particular, I would like to thank Carl Andersen, Alexander Bochman, Jim Delgrande, Ron Fagin, Norman Foo, Amy Greenwald, Benjamin Grosof, Sven Ove Hansson, Andreas Herzig, David Israel, Gerhard Lakemeyer, Bud Mishra, Leora Morgenstern, Abhaya Nayak, Anil Nerode, Maurice Pagnucco, Pavlos Peppas, Don Perlis, Krister Segerberg, Koji Tanaka, David Traum, Johan Van Benthem, Moshe Vardi, Renata Wasserman and Dongmo Zhang. My sincere apologies to anyone that I missed.

I'd like to thank Ananya Mukherjea, whose friendship, love and support made it possible for me to keep at my work right up to the end.

Lastly, I want to thank my brother, Ashutosh (the most important person in the world to me), my *parjaee* Ritu and my nephew Akul for providing a home in India. Their love and affection keeps me going.

Contents

Chapter 1

Belief Revision: The Lay of the Land

1.1 Introduction

The basic plan of this study is as follows: in the first chapter, I survey the fundamental methodology and techniques of belief revision and offer a critique of some of its philosophical foundations. This survey includes examinations of the AGM (or *logic constrained*) method, the methodological constraints on belief revision and the various techniques used for modeling the process of revision. The models of belief revision that I survey are viewed as making *normative* claims and *descriptive* statements about the correct *logic of belief dynamics*. The critique I offer aims at a middle ground: a realistic, descriptively oriented model that maintains the normative model as an ideal and one that admits of a possible implementation. Some key problems with the AGM approach that I consider are problems with the plausibility of the revision postulates as descriptions of reasonable belief change, the inability of the framework to handle *iterated revision* and the allowance of operators that violate intuitions about reasonable belief change.

As a result of this critique, a few desiderata for a plausible model of belief revision emerge: the *distinction between explicit and implicit belief*, the need for *minimal change of beliefs*, the ability to handle iterated (or repeated) revision and the *computational tractability, relevance sensitivity and inconsistency tolerance* of representations of belief and associated revision procedures. In the second chapter, I offer an argument (and amplify those made by others) in favor of a model of belief revision that incorporates relevance sensitivity and inconsistency tolerance. I argue that in order to devise models for belief revision that respect the notion of minimal change, tolerate inconsistency and are computationally tractable, a formal notion of relevance amongst beliefs must be developed and used.

I then move on to consider precedents in the logical tradition for modeling reasoning from inconsistent premises (such as non-adjunctive logics devised by Rescher and Brandom and Schotch and Jennings) and for paraconsistent models of AGM revision. Both of the former approaches use compartmentalization as a technique to quarantine inconsistencies but do not offer a formal definition of relevance. I then examine a model of *localized* belief revision, the Language Splitting model (due to Rohit Parikh) which incorporates a notion of relevance based on the sharing of propositional atoms between formulas.

Finally, I offer two solutions to my critique: the *B-structures* model and the *belief sequences* model. These solutions correspond to the two methodologies present in belief revision: direct revision and logic constrained revision. After discussing their importance, defending their philosophical foundations, and explicating their properties, I offer an assessment of how far these models take us toward more plausible models for belief revision: I evaluate the two models regarding their ranks with respect to the desiderata mentioned previously.

To present a model of belief revision is to make claims about the logic that an agent, thus modeled, employs in inference from its beliefs. Therefore, I make a brief digression into the status of one rule of inference that we lose in the logic of B-structures and belief sequences: the rule of adjunction (for all $\alpha, \beta : \alpha. \beta \vdash \alpha \wedge \beta$). Coupled with this discussion will be a short exploration of how the models presented enable us to deal with the Lottery (due to Henry Kyburg) and Preface (due to David Makinson) paradoxes. I argue that the best way to understand these paradoxes is to view it as an expression of the kinds of inferences that agents might make when forced to consider large bodies of interrelated statements together.

1.2 Notation and Technical Preliminaries

In the following, $\mathcal{L}$ is a finite propositional language with the usual logical connectives ($\neg, \vee, \wedge, \rightarrow, \leftrightarrow$). $\vdash$ denotes derivability, Cn the corresponding propositional closure operation. The constants *true*, *false* are in $\mathcal{L}$. We do not use the symbols ($\top, \bot$) for truth and falsity but reserve them for an alternative interpretation under a four-valued logic. The letter L will often stand for both a set of propositional symbols and for the formulae generated by that set. It will be clear from the context which is meant.

Atoms, Formulae, Theories, and Sequences

- The Greek letters $\alpha, \beta, \gamma \ldots$ denote arbitrary formulae.
- Roman lowercase letters $p, q, r \ldots$ denote propositional atoms.
- Letters K, T denote theories; the letter H (with subscripts for multiple bases) will denote belief bases. $\mathcal{K}$ will denote the set of all theories.

- We reserve the letters σ, τ for belief sequences.

$\alpha \Leftrightarrow \beta$ means that $\alpha \leftrightarrow \beta$ is a tautology, i.e., true under all truth assignments. Similarly, $\alpha \Rightarrow \beta$ means that $\alpha \rightarrow \beta$ is a tautology. The background logic of the belief representation and the associated revision operations is identified with a *consequence operator* Cn. A consequence operator in propositional logic (also termed a *Tarskian consequence operator*) is a function $Cn : 2^{\mathcal{L}} \rightarrow 2^{\mathcal{L}}$ with the following properties (Δ is a set of formulae below):

(i) $\Gamma \subseteq Cn(\Gamma)$ (Inclusion)

(ii) If $\Gamma \subseteq \Delta$, then $Cn(\Gamma) \subseteq Cn(\Delta)$ (Monotonicity)

(iii) $Cn(\Gamma) = Cn(Cn(\Gamma))$ (Idempotency)

(iv) $Cn(\Gamma) = \bigcup\{Cn(\Delta) | \Delta \text{ is finite and } \Delta \subseteq \Gamma\}$ (Compactness)

We will also use the classical notion of entailment $\vdash$ whose properties are indicated below:

- If $\alpha \in X$, then $X \vdash \alpha$.
- $X, \alpha \vdash \beta$ iff $X \vdash \alpha \rightarrow \beta$ (Deduction Theorem)
- If $X, \alpha \vdash \beta$ and $X, \gamma \vdash \beta$ then $X, \alpha \vee \gamma \vdash \beta$ (Disjunction in Premises)

Theories: If X is a set of formulae then $Cn(X)$ is the logical closure of X. In particular, X is a *theory* iff $X = Cn(X)$. A theory K is *finitely axiomatizable* if and only if there is a finite set of formulae Γ such that for any formula $\alpha \in \mathcal{L}$, $K \vdash \alpha$ iff $\Gamma \vdash \alpha$. $K_{\perp}$ represents the inconsistent theory or absurd belief set, or equivalently the entire language $\mathcal{L}$. The size of a formula α (the number of distinct propositional atoms) will be denoted $|\alpha|$.

Binary Relations

Since binary relations ($\mathcal{R} \subseteq \mathcal{D} \times \mathcal{D}$) will often be referred to in the following discussions, we provide a list of common properties[1] (we will use the notation xRy for $(x, y) \in \mathcal{R}$):

- Reflexive: $\forall x \in D. xRx$
- Symmetric: $\forall x, y \in D$, if xRy, then yRx
- Transitive: $\forall x, y \in D$, if xRy, then yRx
- Irreflexive: $\forall x \in D$, $\neg(xRx)$
- Asymmetric: $\forall x, y \in D$, if xRy, then $\neg(yRx)$
- Antisymmetric: $\forall x, y \in D$, if xRy, and yRx then $x = y$
- Connected: $\forall x, y \in D$, either xRy or yRx
- Equivalence: R is Reflexive, Symmetric and Transitive
- Preorder: R is Reflexive and Transitive
- Partial Order: R is Reflexive, Transitive and Antisymmetric
- Total Order: R is a Partial Order and Connected
- Simple Order: R is Transitive and Antisymmetric
- Strict Partial Order: R is Asymmetric and Transitive
- Strict Simple Order: R is Asymmetric, Transitive and Connected

[1] I owe the list below to Maurice Pagnucco.

1.3 Models of Belief Revision and Rationality

The study of *belief revision* is central to any effort to formalize intelligent behavior. As an agent (human or otherwise) lives out its life, it must deal with the new information it receives in some principled fashion. Often such information will be useless. false or confusing. often this information will be felt to be worth accepting. The agent's different orientations with respect to the information it receives, its principled reasons for changing its beliefs in response to this new information, and its strategies for changing its beliefs are notions that must be understood if a sound theory of rationality and reasoning are to be developed for it. Such a theory for the agent in question will be a mix of normative claims and descriptive statements. The normative claims will rely on intuitions about what rational reasoning and belief change *should* look like. the descriptions on facts about how agents *actually* change their beliefs and reason from them. Central to such theories are several concepts: the procedure by which beliefs come to be formed. the criteria for the acceptance of a belief. the reasons and strategies adopted for change of belief and the kinds of inferences an agent makes from its beliefs. Belief formation, retention, and change, then, are the crucial notions in such a model of rationality and reasoning.

In this study, we will have the opportunity to examine the dynamics of belief i.e., the change of beliefs of an intelligent agent over time. An agent in this study can be a human being or a reasonable simulation of it, such as a highly autonomous software program or rules-based system. Whatever the form of the agent. the formalization of the belief dynamics of the agent will significantly inform the models of rationality that we construct for it[2].

[2]Should we also be concerned with formalizing changes in desires? In this study, I follow the line taken by Harman [53] in that I do not consider changes in desires as part of reasoning. While

Conversely, our pre-theoretic intuitions of rationality and plausible reasoning will inform the models of belief revision that we construct. Belief revision models deal with how we change our beliefs in the light of new information and how we choose to store and organize our beliefs so that we expend the least cognitive labor and retain a measure of coherence while revising our beliefs. If inquiry is understood as the truth-directed, always-curious, rational transformation of beliefs over time[3], then an adequate grasp of what constitutes rational change of belief and how it is best accomplished is crucial to any model of belief revision and practical reasoning.

The study and formalization of models of rationality, reasoning and belief revision is an enterprise with each of its components informing our understanding of the other and each in turn being reliant upon others for their own clarification. Each point of entry into this enterprise is as significant as the other. In this study, through an investigation of belief change, I hope to shed some light on models of rationality and plausible reasoning.

Logical studies of belief change have taken as a guiding light, the model developed by Alchourron, Gärdenfors and Makinson [2, 34]: the AGM model. My discussion below assumes this model as fundamental. It is not the only method available: I defer a description of an alternative approach till the concepts necessary to understand the discourse in this area have been presented.

the study of belief-desires goes together in that beliefs are dispositions to assert to statements or dispositions to act in certain ways given certain desires, the lack of coherence conditions on desire such as consistency makes it impossible to formalize or treat the two together in similar fashion. That is, it is not clear what, if any form, a normative theory of desires could take. Desires, then, are not part of reasoning, but simply result from reasoning.

[3]Stalnaker, [99].

1.4 Methodological Factors in Belief Revision

Epistemic States

The most fundamental concept is that of an *epistemic state.* What best represents an agent's *doxastic commitments*[4]? What *representation* of the epistemic state is the most perspicuous? We begin with an idealization of epistemic states used in the AGM model.

A. Consistency: No inconsistencies are to be tolerated in a *coherent* epistemic state[5].

B. Logical Closure: An agent's beliefs should be closed under logical implication i.e.. the agent's *doxastic commitments are consistently completable.* An agent's beliefs have not been adequately represented if there is a proposition logically implied by them which it does not already believe[6]. In that case. the agent has the choice of adding the implied proposition to its beliefs or giving up one or more of the implying beliefs.

The conditions above entail the vacuity of inconsistent belief states: a contradiction entails everything (by *ex falso quodlibet*) and so by closure, an agent's inconsistent belief state contains all beliefs. The problem with a belief state in which all proposi-

[4]This notion is due to Isaac Levi who uses it to establish the doxastic commitment versus doxastic performance distinction: the agent is normatively committed to consistency and deductive closure in its beliefs while its actual performance may differ due to human limitations. We return to this distinction in the next chapter.

[5]Allied to this condition are requirements that a particular representation of beliefs cannot be a belief system if it is not a self-correcting system i.e., a system that incorporates a way of changing in response to new information that increases the probability of propositions in the system being true. A representational system which has the capacity to entertain both a proposition and its negation is not self-correcting under this view and hence cannot be a belief system: see Bernard Williams [107] and Michael Pendelbury[81].

[6]If you believe α. and if you believe β. then you believe $\alpha \wedge \beta$. (the rule of inference used here is: $\alpha, \beta \vdash \alpha \wedge \beta$). Or, if you believe $\alpha \rightarrow \beta$, and if you believe α. then you believe β (the rule of inference here is *modus ponens*).

tions are believed is that it is utterly useless:

> No belief state in which all propositions were believed could distinguish any actions as appropriate or inappropriate.[7]

Ellis [21] places differently phrased requirements on *rational belief systems* (while noting that "strictly rational belief systems are only for the gods"). One is the notion of *strict rationality* which corresponds to the consistency condition and the other is the notion of *completability* which corresponds to deductive closure. Gärdenfors [34] has shown that Ellis' requirements (and those of the traditional conception) are met by a *belief set* (i.e., a *theory* closed under consequence) and offers the following formalization:

DEFINITION 1 *A set K of sentences is a (non-absurd) belief set iff (i) $\bot$ is not a logical consequence of the sentences of K and (ii) if $K \vdash \beta$, then $\beta \in K$.*

Epistemic Attitudes

Given an epistemic state, we need to understand the orientation of the agent with respect to new information: its *epistemic attitudes*. There are three basic attitudes that we consider in this study. Assume a closed and consistent set of beliefs K and a new proposition α: the agent's attitudes are as follows:

- The proposition α can be *accepted*: $\alpha \in K$.
- The proposition α can be *rejected*: $\neg\alpha \in K$.
- The proposition α can be kept in *suspense* or *indetermined*: $\alpha \notin K$ and $\neg\alpha \notin K$.

[7] Stalnaker [99], page 83.

The basic operations in the AGM model of belief revision consist of attempts to formalize these attitudes[8]. Classifying epistemic attitudes as above and choosing the representation of beliefs as sets of sentences with no additional structure makes it clear that the AGM model of belief revision does not possess a notion of *degrees of belief*[9].

It might be felt that the model for epistemic states and attitudes presented above presumes a strong connection or special relevance of classical deductive logic to reasoning and belief change. There is no reason, though, to believe that such a logic alone will *decide* what *strategy* one would follow in revising our beliefs since beliefs can often be revised in epistemically different, logically equivalent ways. These revisions can be logically equivalent while being different in that the end result (a closed belief set) is the same while their content (the propositions believed) is very different. The central problem of belief revision is that logical considerations *alone* do not tell us how to revise our beliefs. As an example, consider the following situation:

Example: The case of the twin cities: I believe that Delhi has hot summers and Tashkent has cool summers. Now, I am told that the two cities' climates are indistinguishable with respect to their summers. Which one of my beliefs should I give up? Dropping either one of the two would seem to do the trick of preserving logical consistency. In addition, I would perhaps need to drop the belief that Tashkent is a nicer place to visit in the summers than Delhi. Perhaps, I should drop the one that I believe '*less*'[10]? Perhaps, the information about Tashkent comes from

[8]Generalizations of the formal machinery have been extended to the case when epistemic inputs are sets of sentences: see [49, 32, 45, 23].

[9]Unless an ordering of epistemic entrenchment is added to a belief set. See Appendix A.

[10]At the moment, the concept of believing one proposition 'more' than the other is undefined, so I will leave this open to any intuitions that the reader might have about the relative strength of beliefs.

testimony while the information about Delhi comes from bitter experience of many a hot summer? These are not choices guided by logic. These are guided by extra-logical factors. Matters get more complicated if we consider possible the absence of the original reasons for our beliefs[11]. In the end, I will probably drop the belief that costs me the least, cognitively[12], to drop. I do follow logical considerations in obeying a certain integrity constraint, in that my choices will be restricted to those that make my beliefs consistent. Logic alone, though, does not help me in making up my mind: the logical structure of our beliefs cannot *decide* how our beliefs would be or should be revised. No matter what the representation of beliefs, much more needs to be said about the *structure* of an epistemic state before anything can be said about revision strategies[13].

The Principle of Categorial Matching

The *principle of categorial matching*[14] says that the structure of an epistemic state should be stable under revisions. So, the end result of revising a belief set should be another belief set. Since our strategies for revision are tied into the way we chose to represent our beliefs[15], a revision strategy that changes the representation of beliefs on its application would necessitate the use of a new strategy for the new representation created by it. The problem of *iterated belief revision* (the revision of a belief set K by α and then the revised set $K * \alpha$ by a new proposition β) then becomes especially acute since the strategy in question can only be applied once.

[11]Harman, [53].

[12]Perhaps in terms of informational value or some other feature.

[13]Stalnaker [99], page 97.

[14]Formulated in [37, 14].

[15]In [30] Friedman and Halpern suggest the need for an *ontology* of belief representation. What they call rather confusingly "an ontology" is however, just a nod in the direction that the choice of representation of the beliefs of an agent is crucial to the understanding of how those beliefs are to be revised.

The Minimal Change Requirement

Agents operate with a certain inertia in their beliefs as a strategy to reduce cognitive stress (a fundamental rationality constraint on belief revision). This intuition is captured by the *minimal change* requirement[16]:

> **The Principle of Minimal Change:** After a change of epistemic states, the new state should be as *similar* as possible to the old state.

We can think of 'similarity' as a measure of *how many of the old beliefs are retained*[17] or *how much of the content of the old beliefs is retained.* It is not clear which of these two notions is intended. The best way to do justice to intuitions about similarity is to interpret these principles charitably and take them to be indicating minimal change of quantity *and* content of beliefs.

Such a condition is formalized by what Gärdenfors has called the *preservation criterion*[18]:

> If a belief state K is revised by a sentence α, then all sentences in K that are independent of the validity of α, should be retained in the revised state of belief K'.

One corollary of this requirement is that the relevant notion of minimality will be dependent upon which notion of epistemic states is used in our representation. Examples have been provided by Hansson [47] to show that taking belief sets to be logically closed can often lead us to delete more beliefs than are actually warranted by a belief change operation. Minimal change is not *always* the best description of a

[16]This can be traced back to Quine's notion of changes being kept as small as possible in the *Web of Belief* in "Two Dogmas of Empiricism" [86].

[17]As Harman does with the following **Principle of Conservativity:** When changing beliefs in response to new evidence, you should continue to believe as many of the old beliefs as possible. Rohit Parikh and Henry Kyburg both point out that this condition makes sense only in the context of finite sets of sentences.

[18]Gärdenfors, [33].

particular belief change: we can imagine situations in which a particular belief causes a radical enough reexamination of our epistemic states such that we lose belief in a great many propositions that we formerly believed[19].

The Foundationalist versus the Coherentist Approach

The next fundamental concept is to clarify the *status* of beliefs in the agent's belief set i.e., the standing of individual beliefs in an agent's belief set and the difference between *foundational* and *coherentist* approaches. Foundational and coherentist views of belief are theories about how our beliefs are structured and what the nature of justifications for our belief amounts to. Harman [53] further makes a distinction between foundationalist and coherentist theories of justification and those of belief revision: an analysis of the kinds of justifications our beliefs possess guides our strategy for the revision of those beliefs. Harman suggests that the primary issue in belief change is not so much the nature of justifications for an agent's beliefs but whether the agent needs to hold on to those justifications.

The foundationalist approach presupposes a special class of beliefs termed *basic* beliefs. Beliefs in a foundational system rely for their justification on other beliefs which in turn rely on others, terminating in a set of beliefs that may be thought of as self-justifying. The belief base approach (described below) with its emphasis on explicit and derived beliefs is foundationalist in nature: the criterion for a particular proposition to be believed by some agent is not that the proposition is a member of a particular belief set, but that the proposition is a logical consequence of some set of basic beliefs. In a foundationalist model of belief revision we first drop those beliefs that do not currently have satisfactory justification and then add those beliefs that either need minimal justification or are justified on the basis of beliefs already present

[19]Kuhnian paradigm shifts in science could be an example of this kind of change.

in the agent's belief set.

The coherentist approach denies the existence of any special set of beliefs: all beliefs have the same status. Beliefs are justified in the way they interact or *cohere* with other beliefs, and it is the relationship with other beliefs that determines whether a belief is justified or not. The AGM approach, with its insistence on logical closure, consistency and the minimal change principle, is coherentist in nature: no beliefs in the belief set possess special status and changes are made so as to increase and preserve overall coherence. There is no demarcation in the belief set in terms of relative justification or importance of beliefs: no special justification is required for currently held beliefs[20].

Harman [53] suggests that while foundationalist theories of belief revision seem to carry greater normative import, in actual practice, agents will revise their theories in line with what the coherence theory suggests. The primary argument against the foundationalist theory is that agents often *do not retain their original justifications for a belief* as a time and memory saving strategy for avoiding 'clutter' and for explaining cases[21] that test our intuitions on plausible belief change. Agents tend to hold on to beliefs even when the justifications for those beliefs have been lost: the beliefs held on to have value despite such a lack of justification. However, the suggestion (as in the coherentist model) that there is a holistic assessment of new beliefs seems implausible: we will contest this claim later in more detail.

Kinds of Revision

Gärdenfors and Rott [37] make a distinction between two kinds of methodologies for modeling revision, *direct belief revision* and *logic constrained revision.*

[20]The *B*-structures model that we present in Chapter 5 can be viewed as a combination of both foundationalist and coherentist views.
[21]Harman, [53], pp 32-42.

1. Direct Belief Revision

Given a particular representation of beliefs, this approach inserts and deletes beliefs *without conforming to any integrity constraints such as closure and consistency.* Typically, this approach uses trivial change operations and provides a companion method of inference from the chosen representation of beliefs that is *either paraconsistent or non-monotonic.*

Gärdenfors and Rott describe this approach as reducing the "dynamics of belief states to the statics of inference relations"[22]. By definition, the direct method is *constructive*; the actual content of the agent's epistemic state is given by the inference operation so defined since the epistemic states just *are* the non-monotonic consequences of the inference operation in question. Examples of the direct revision approach can be found in the *preferred subtheories* approach of Brewka [8], the *ordered theory presentations* approach of Ryan [95] and in the *truth maintenance system* of Doyle [17].

2. Logic Constrained Revision

This method (the AGM approach) uses the integrity constraints contained in the classical notion of an epistemic state (defined above) as constraints for the process of belief change. Taking the integrity constraints as guides renders the change operation into one requiring considerable ingenuity since the construction of a theory after removing beliefs and the choice of beliefs to be dropped are non-trivial tasks. The advantage of this approach is that classical propositional logic can be used as the underlying logic for all change operations. If preserving propositional logic as the correct logic of belief change is a motivation, then this approach offers a *prima*

[22]This approach is also termed the *vertical* perspective by Rott [94] (when dealing with a belief base).

facie advantage[23]. Under the logic constrained or AGM approach we can isolate the following assumptions:

- The agent holds beliefs that constitute some logically closed theory *and that is all there is say about the agent's epistemic state.* There is no more structure: the theory is 'flat' and undemarcated (since it is the coherentist view).
- The agent's knowledge is changing but the world remains static. This is the requirement of a constant language which corresponds to the lack of *conceptual expansion* on part of the agent. Given a language in which to describe the world. the agent can only modify its belief set by beliefs expressed in that language.
- Maximal inertia under belief change is desired.

1.5 Update Operations and the AGM Axioms

If K is a set of beliefs, and a new piece of information α is received, how do we revise our beliefs? If α is consistent with K then we just add α and close under consequence. Matters get more complicated when α is not consistent with our existing set of beliefs. In the logic constrained method there are three 'basic' operations:

Expansion: A new belief is added to the belief set with no attention paid to the consequences of the set thus formed. The belief set that results from expanding a belief set K by a new input α is denoted $K^{\dot{+}}_{\alpha}$; the old set K is contained in $K^{\dot{+}}_{\alpha}$.

Revision: A new input that is typically inconsistent with the old belief set is added to the belief set. In order to maintain the integrity constraints. the belief set is revised

[23]This approach is also termed the *horizontal* perspective by Rott [94] (when dealing with a belief base).

to accommodate this new belief; some previously held beliefs must be dropped. The result of revising a belief set K by a new input α is denoted K^*_α.

Contraction: Some beliefs are dropped from the belief set. In order to satisfy integrity constraints, other beliefs must then be given up. As an example, consider the epistemic state consisting of $\alpha, \beta, \beta \rightarrow \gamma, \alpha \wedge \gamma \rightarrow \delta$. If we decide to drop δ from this state, we must give up one of the first three propositions (as well) in order to maintain the logical closure of the epistemic state. The result of contracting K by α will be denoted K^-_α.

Since there is no logical way of deciding between competing alternatives for a belief revision or contraction operation, a *non-constructive* approach is to present *operators* for each of the above processes and then present postulates (also termed *rationality postulates*) or axioms for these operators. The AGM axioms for the update operators are useful in telling us what logical properties an update operator should satisfy. This establishes the distinction between constructive and non-constructive methods of belief revision. Axiomatized or non-constructive approaches simply constrain the *logical form* of the change of the belief set, and do not specify its actual content after revision. Constructive methods provide us with operations that inform us of what beliefs are actually held by the agent in the new epistemic state.

Revision operators can be defined as follows:

DEFINITION 2 *A revision operator* $(*, \dot{+}, \dot{-})$ *is a function from the belief state and the relevant input or subject formula to a new belief state:*$\dot{+}, *, \dot{-}$: $\mathcal{K} \times \mathcal{L} \rightarrow \mathcal{K}$.

A revision operator takes as input a theory and a formula and maps it onto a new theory. Less formally, a revision operator encapsulates the agent's epistemic strategy vis-a-vis new information. In the logic constrained approach, the tacit assumption is that all information about the revision is contained in the revision operator.

In the following sections, I briefly present the rationality postulates for the operations mentioned above (in the discussions that follow, I will refer to the axioms by both their numeric identifier and their formal name).

Expansion

Axioms for expansion operators ($\dot{+}$: $\mathcal{K} \times \mathcal{L} \rightarrow \mathcal{K}$) are given below[24]:

($K^{\dot{+}}$ 1). $\forall K, \alpha, K^{\dot{+}}_{\alpha}$ is a theory. (Closure)

($K^{\dot{+}}$ 2). $\alpha \in K^{\dot{+}}_{\alpha}$ (Success)

($K^{\dot{+}}$ 3). $K \subseteq K^{\dot{+}}_{\alpha}$ (Inclusion)

($K^{\dot{+}}$ 4). If $\alpha \in K$ then $K = K^{\dot{+}}_{\alpha}$ (Vacuity)

($K^{\dot{+}}$ 5). If $K \subseteq T$, then $K^{\dot{+}}_{\alpha} \subseteq T^{\dot{+}}_{\alpha}$ (Monotonicity)

($K^{\dot{+}}$ 6). $\forall K, \alpha, K^{\dot{+}}_{\alpha}$ is the smallest belief set that satisfies $(K^{\dot{+}}1) - (K^{\dot{+}}5)$. (Minimality)

Justification for the axioms: Closure expresses the intuition that $\dot{+}$ takes a belief set and a sentence as input and produces a belief set as output. Success says that the new input is accepted in the new epistemic state. Inclusion says that no beliefs are retracted in the process of expansion (i.e., there is full inertia of beliefs under expansion). Vacuity has the seemingly trivial consequence that no expansion is necessary if the input is already believed. Monotonicity says that if one belief state contains at least the same amount of information as another, then the expansion of that belief set will contain at least as much information as the expansion of the other with respect to the same epistemic input. Minimality is a concise expression of the Principle of Minimal Change applied to the addition of new beliefs to an

[24]It will be noted (once axioms for revision and contraction have been described) that there is no counterpart to the Extensionality axiom in expansion, but a plausible case could be made for its inclusion in the axioms above.

epistemic state: the smallest change possible to an existing belief state is made in order to accommodate the new information. The axioms above and the operation of expansion are linked by the following representation theorem:

THEOREM 1.1 [2] *The expansion operator* $\dot{+}$ *satisfies* $K^{+}1 - K^{+}6$ *iff* $K^{\dot{+}}_{\alpha} = Cn(K \cup \{\alpha\})$.

Contraction

Contraction ($\dot{-}: \mathcal{K} \times \mathcal{L} \rightarrow \mathcal{K}$) is guided by the following postulates:

(K^{-} 1). $K^{\dot{-}}_{\alpha}$, the contraction of K by α, is a theory. (Closure)

(K^{-} 2). $K^{\dot{-}}_{\alpha} \subseteq K$ (Inclusion)

(K^{-} 3). If $\alpha \notin K$, then $K^{\dot{-}}_{\alpha} \subseteq K$ (Vacuity)

(K^{-} 4). If $\not\vdash \alpha$, then $\alpha \notin K^{\dot{-}}_{\alpha}$ (Success)

(K^{-} 5). If $\alpha \in K$, $K \subseteq (K^{\dot{-}}_{\alpha})^{\dot{+}}_{\alpha}$ (Recovery)

(K^{-} 6). If $\alpha \Leftrightarrow \beta$, $K^{\dot{-}}_{\alpha} = K^{\dot{-}}_{\beta}$ (Extensionality)

(K^{-} 7). $K^{\dot{-}}_{\alpha} \cap K^{\dot{-}}_{\beta} \subseteq K^{\dot{-}}_{\alpha\wedge\beta}$ (Intersection)

(K^{-} 8). If $\alpha \notin K^{\dot{-}}_{\alpha\wedge\beta}$, then $K^{\dot{-}}_{\alpha\wedge\beta} \subseteq K^{\dot{-}}_{\alpha}$ (Conjunction)

Inclusion says that no new beliefs should be introduced into the new epistemic state after contraction. Vacuity states that no action is necessary if the epistemic input is not accepted (a consequence of the Principle of Minimal Change). Success says that if it is *possible* to remove the input in question, then it will be retracted from the current state (this possibility is to be understood as either logical or physical possibility). The only kind of input which cannot be retracted is a logical truth, for as before, logical truths must be included in all possible epistemic states. Recovery states that if we were to retract a belief from K and then expand by the same formula, the resultant epistemic state would be identical to the state before the contraction and expansion. This, once again, is an expression of the Principle of Minimal Change[25]. Extensionality expresses the principle of the *irrelevance of syntax* i.e., logically equivalent formulas should have the same effect on the epistemic state. Intersection says that if one does not give up belief in γ when giving up belief in α nor in giving up belief in β, then one should not give up belief in γ when giving up belief in the conjunction $\alpha \wedge \beta$. Conjunction states that, if one were to give up α when giving up the conjunction $\alpha \wedge \beta$, then whatever beliefs one gives up in contracting by α, should also be given up in contracting by $\alpha \wedge \beta$.

Revision

Revision ($* : \mathcal{K} \times \mathcal{L} \rightarrow \mathcal{L}$) is axiomatized as follows:

(K^* 1). K^*_α, the revision of K by α, is a theory. (Closure)

(K^* 2). $\alpha \in K^*_\alpha$ (Success)

(K^* 3). $K^*_\alpha \subseteq K^+_\alpha$ (Inclusion)

(K^* 4). If α is consistent with K i.e., if $\neg\alpha \notin K, K^*_\alpha = K^+_\alpha$ (Preservation)

(K^* 5). K^*_α is consistent if α is, i.e., $K^*_\alpha = K_\perp$ iff $\vdash \neg\alpha$ (Vacuity)

(K^* 6). If $\alpha \Leftrightarrow \beta$ $K^*_\alpha = K^*_\beta$ (Extensionality)

(K^* 7). $K^*_{\alpha\wedge\beta} \subseteq (K^*_\alpha)^+_\beta$ (Superexpansion)

(K^* 8). If $K^*_\alpha \nvdash \neg\beta$, then $(K^*_\alpha)^+_\beta \subseteq K^*_{\alpha\wedge\beta}$ (Subexpansion)

Closure is familiar to us by now: the result of the revision should be another theory. Success states that the new information should be included in the new epistemic state (this axiom can also be thought as saying that the revision should be *effective*). Inclusion says that expansion always gives a 'larger' belief set when incorporating new beliefs as compared to revision (for an epistemic input α, if $\neg\alpha$ is already ac-

[25]See Hansson [46] and Levi [66] for a discussion of the Recovery postulate and for suggestions that it not be considered a valid postulate for contraction operations.

cepted, expansion will simply result in the set $K_\perp$ and this postulate holds trivially). Preservation expresses the intuition that if the negation of the epistemic input is not accepted, revision reduces to expansion. Vacuity (in its original form) says that the only situation in which a revision creates an inconsistent epistemic state is when an agent accepts logically contradictory information. Extensionality, as above, is an expression of the principle of *irrelevance of syntax*. The last two postulates are supplementary and are often regarded as generalizations of Inclusion and Preservation. Super-expansion says that any belief included in the revision of K by $\alpha \wedge \beta$ should also be included if we initially revise K by α and then expand by β. Sub-expansion says that if β is not rejected in revising K by α, then any belief included by first revising K by α and then expanding by β should also be included in the revision of K by $\alpha \wedge \beta$. This last condition renders the sets $K^*_{\alpha\wedge\beta}$ and $(K^*_\alpha)^+_\beta$ equal.

The Levi and Harper Identities: Interdefinability of Operators

The revision operators discussed above are interdefinable by the Levi and Harper identities. The *Levi identity* [64] defines the revision operator in terms of the contraction and expansion operator. It expresses the intuition that in order to revise by a new input, the agent must first contract with the negation of the information to a new belief state, and then expand in order to include the new piece of information. In fact, Levi insists that all belief change be thought of as consisting only of expansions and contractions.

$$K^*_\alpha = (K^-_{\neg\alpha})^+_\alpha$$

So, if the new information is consistent, we expand K, if not, we contract K to $K' = K \dot{-} \neg\alpha$ and then add α to K. The following representation theorem establishes the relationship:

THEOREM 1.2 [2] *Let $\dot{-}$ be a contraction function satisfying postulates* $(K^{\dot{-}}1) - (K^{\dot{-}}4)$ *and* $(K^{\dot{-}}6)$ *and* $\dot{+}$ *the expansion function satisfying postulates* $(K^{\dot{+}}1) - (K^{\dot{+}}6)$. *Then the revision function* $*$ *obtained from the Levi Identity satisfies* $(K^{\dot{+}}1) - (K^{\dot{+}}6)$. *Moreover, if* $\dot{-}$ *satisfies* $(K^{\dot{-}}7)$ *then* $*$ *satisfies* $(K^{*}7)$ *and if* $\dot{-}$ *satisfies* $(K^{\dot{-}}8)$, *then* $*$ *satisfies* $(K^{*}8)$.

The *Harper identity* [54] defines the contraction operator in terms of the revision operator and set intersection:

$$K_{\alpha}^{-} = K \cap K_{\neg\alpha}^{*}$$

Contracting K by α consists of those beliefs in K that are retained when revising K by $\neg\alpha$: since $K_{\neg\alpha}^{*}$ represents the minimal change required to (consistently) integrate $\neg\alpha$, it should contain a large part of K that does not imply α.

1.6 Problems with AGM Axioms and Update Operations

Some powerful representation theorems (described in Appendix A) for AGM models of belief revision can be obtained. These characterize a set of operations on belief sets that obey all the integrity constraints of AGM revision and conversely, AGM postulates are shown to be exactly those that conform to these constructions. Clearly, the framework provided by the AGM approach lends itself to the construction of a beautiful formal machinery. The question of its plausibility, however, can be raised on several dimensions: the assumptions underlying the basic operations on belief change themselves, the plausibility of the postulates and the inability of some of the constructions to handle iterated belief revision. To illustrate some of these problems, I describe an example (which will be referred to later on as well in this study when our new models have been presented).

Example: A fighter pilot arrives for work, is briefed for patrol and told that according to the latest radar reports, enemy aircraft are in the region. The pilot takes off believing that enemy aircraft are present in his sector. The pilot then observes an aircraft on the horizon and classifies it as an enemy aircraft. He seems to be justified in his new belief: he is an experienced pilot and has received many hours training in aircraft recognition. The pilot sets up his weapon systems as gets closer and then observes that it is a two-seater aircraft. Now the pilot believes that he is closing in on a hostile, two-seater aircraft. As he gets even closer, he notices the markings on the fuselage and realizes it is an aircraft of his own airforce. The pilot turns off all weapons and sends a friendly radio signal to the aircraft in front of him. A few minutes later, the pilot receives a radio message from base indicating that all enemy aircraft have left the area. The pilot has undergone several changes of his belief states in the course of his activities. As we see below, some AGM axioms and AGM-compatible updates provide implausible descriptions of these changes.

Problems with Contraction

We begin with contraction. The AGM model does not explain *why a rational agent would stop believing in one of its currently held beliefs*: we have a non-reason based approach to defining contraction. However, agents have reasons for their beliefs (even if they don't retain them over time) and therefore, the operation of contraction should only happen in response to some new information being received. Contraction is more plausibly thought of as *part* of the revision operation: the agent contracts its beliefs when it receives information that it wants to integrate into its epistemic state.

If I believe α, one way to think of me as contracting by α is to think of me having been told $\neg\alpha$ by a reliable source.

Examples:

- If I believe that the Mets won last night because they were leading when I went to sleep, and if I decide to drop that belief in the morning, it's because I have found out that they lost. I revise by "The Mets lost". I do not just drop my belief in "The Mets won". It's possible that I find out the radio station I was listening to at night was just playing a cruel trick on Mets fans and not believe "The Mets won" any more. The reason I stop believing this proposition is because I have received information that makes me do so: I do not just drop the old belief.

- Our pilot above does not just lose the belief that the aircraft in front of him is a hostile one; he does so because he has received explicit information (a clear visual sighting of markings on the fuselage) that it is a friendly craft. He loses his belief that there are enemy aircraft in the area after receiving explicit information to that effect from ground radar controllers.

This criticism applies to the Levi identity as well in its assignment of a primary position to contraction. A more plausible description of an agent's epistemic strategy is that the agent commences the revision process by first believing the new piece of information and then weeds out inconsistencies as and when it is possible and convenient to do so. In the case of the pilot, he realizes that the aircraft is not an enemy aircraft and in doing so, comes to believe *first* a piece of information that is inconsistent with his older belief that the plane in front him is an enemy craft. Suppose now that the pilot has to report back to his home base on any sightings. It is at this stage that the pilot reports that he has not seen any enemy aircraft yet. The pilot does not commence the revision process by dropping the belief that there is an enemy plane in the area, he initiates the process by coming to believe that there is an enemy aircraft and then revising his older stock of beliefs. If so. why not consider revision as expansion followed by contraction? The AGM model rejects this notion since expansion to an inconsistent state means representing the agent's belief state by $K_\perp$, the trivial belief set. Hansson [49] suggests using a belief representation that can coherently represent inconsistent epistemic states i.e., logically open belief bases (we return to this issue below) and *reversing* the Levi identity.

Lehmann [63] argues that there should be two sets of postulates for contraction. Suppose an agent learns the formula $\alpha \wedge \beta \ldots \wedge \zeta$ and then its negation $\neg\alpha \vee \neg\beta \ldots \vee \neg\zeta$. The first information is incorrect. but it could be almost correct (the reliability of the source of information is still in question). One could believe then. that only a few components of the conjunction are false while others are still true (the falsity of one, obviously. being enough to render the conjunction false). This suggests that there are two kinds of contractions. In the first kind. one drops a particular belief since the source of the belief has been shown to be unreliable. We lose full faith in the proposition (and in others received from the same source) since the source of the information itself has been compromised. In the second kind, we retract a proposition because we receive information that contradicts it. In this case, we might still retain partial belief in the proposition.

An example of a severe contraction upon being informed of the lack of reliability of a source of information: consider the pilot above on patrol at dusk who observes the (potentially hostile) aircraft in poor and fading light. He comes to the conclusion that the aircraft is hostile and a two-seater. Upon realizing (as he gets closer) that the aircraft is not hostile after all, the pilot now revises *all* beliefs formed in that poor and fading light.

Problems with the Postulates

The AGM postulates are intended as rationality postulates. That is, any revision operator that satisfies the postulates is claimed to be *fully* 'rational'. However, it is debatable whether some of these operators are indeed 'rational': the AGM operators are more plausibly viewed as only necessary but not sufficient for rational revision operators. Ryan [95] has contended that the AGM axioms be considered neither '*sound*' nor '*complete*' with respect to intuitively rational belief revision. A lack of 'soundness' in this regard means a failure to conform to intuitions about plausible belief revision. The failure of 'completeness' corresponds to admitting operators that violate fundamental methodological constraints on rational belief revision. It is to this problem that we first turn.

The Trivial Update

The AGM axioms possess the unfortunate property of being consistent with the *trivial update*:

DEFINITION 3 *If α is consistent with K, then $K * \alpha = K \dot{+} \alpha$, otherwise $K * \alpha = Cn(\alpha)$.*

That is, if α is inconsistent with K, all the old information in K is dropped. This sort of change is extremely drastic: the acceptance of new information inconsistent with the agent's epistemic state causes the agent to wipe its slate clean and begin afresh. This does not correspond to any sort of reasonable model of belief change; it's not the sort of thing we do when we receive information that contradicts something that we know. For example, should the pilot in the example above also drop his beliefs about the color and seating capacity of the aircraft once he finds out that it is not a hostile entity? The only situations in which one could imagine the trivial update would be the receipt of some extremely fundamental piece of information that has the effect of wiping out everything that I believed. For example, were I to receive information today that I was a brain-in-a-vat, it is likely that I would withhold belief in every one of my beliefs since all of them would have been severely compromised. Furthermore, were I to receive information that I was a brain-in-a-vat, my reaction to this would be a gradual suspicion of all that I knew while I went about trying to confirm this crucial piece of information. The infection of the entire belief set with the new inconsistency and the radical change it precipitates are pointers to a less drastic way to handle inconsistency and a better way *to only change the part of the theory that is actually relevant to, or affected by, new information*[26].

In the case of the trivial update, the most pertinent questions are: surely the information α only renders *part* of our belief set inconsistent? Why is an update allowed by the axioms that drops *all* old beliefs upon receipt of a single piece of information that conflicts with them? A more realistic model would do better to block such a radical operation from ever taking place unless and until some fairly strong conditions had been met.

The Supplemental Axioms

The AGM supplemental axioms (K^* 7), (K^* 8) are intended as approximations of iterated revision. They do not contain expressions like $K * \alpha * \beta$ and therefore, cannot be thought of as placing any explicit constraints on iterated revision. Coun-

[26]In Chapter 5 I will discuss the Language Splitting model [80], explicitly intended to provide update operators that are *consistent* with the **AGM** axioms and which *block* the trivial update.

terexamples can be constructed to show that these axioms are not plausible axioms for belief revision. The problem lies in an equation of two epistemic states: the state corresponding to receipt of information α, β and the state corresponding to the receipt of the information $\alpha \wedge \beta$. The receipt of the conjunction of two propositions, however, results in a different epistemic state than on the receipt of just the conjuncts since the conjunction lets us draw inferences not available on receiving each conjunct separately. Besides learning that the two propositions can be true together, each proposition has an impact on the epistemic state that is not captured by their sum. This is what equating them to their conjunction does: the whole, though, is greater than the sum of its parts.

Take, for instance, $\alpha = \neg p \vee q$ and $\beta = p \vee q$. Then $\alpha \wedge \beta$ is equivalent to q and says nothing about p, i.e., learning the conjunction is the same as learning the proposition q by itself. Revising by $\alpha \wedge \beta$ is not the same as revising first by α and then β since revising by α would cause me to drop some p beliefs whereas revising by $\alpha \wedge \beta$ will not have this effect. In the first case, I learn that either p is not true or q is, and in the second case, I learn that either p is true or q is. When I learn the conjunction of the two however, I learn that q is true and learn nothing about the connection of p to q.

More informally, consider the following : I might believe quite firmly that my brother is not taking yoga classes ($\neg\beta$); he just is not the type. I also believe that if my cousin has a frequent cough, he has asthma ($\gamma \rightarrow \alpha$); all the evidence, in my opinion, points to that diagnosis. Now, I am told that $\neg\gamma \vee \beta$ i.e., that either my cousin does not have a cough or my brother is taking yoga classes. I conclude that my cousin does not have asthma. Then I am told $\gamma \vee \beta$, that my cousin has a cough or my brother is taking yoga classes. Now, I conclude that my cousin has asthma. However, I *never* acquired the belief that my brother is taking yoga classes. I would have, though, had I been told $(\neg\gamma \vee \beta) \wedge (\gamma \vee \beta)$.

The Postulate of Vacuity

The postulate of Vacuity is problematic in that it asks us to accept the notion of an agent accepting an *arbitrary formula*: it suggests that an agent can accept logically contradictory information or revise by the constant *false*. Parikh [80] has argued that this makes no intuitive sense: what is it to *learn* contradictory information? It would be far more reasonable to think of postulates that block revision by the constant $false$ but accept revisions by *true* or the tautologies. We are often provided tautological information ("either your mother is at home or she is not!") so the possibility of receiving and accepting tautological information should not be blocked by a model of belief change. The result of revising by $true$ should just be the original belief set, since tautologies bring about no change in our epistemic state. Modifying the axioms lets us treat revision by $true$ as an *identity operator*: $K * \mathit{true} = K$ always. We can then assume that both K and α are *individually* consistent while their union might not be.

Two "uncontroversial" Postulates

- $\alpha \in K * \alpha$

The problems with this postulate are the assignment of a special status to new information[27], and more importantly, *its immediate loss of that special status on*

[27] It is perhaps unfair to say that this is a problem since in the AGM model it is assumed that new information is always accepted i.e., "new information" should really be read "new information which an agent has decided to accept". Having defined it this way, it is tautologous to say that $\alpha \in K * \alpha$. In an alternative view, the mere presentation of new information to an agent can have an effect on its epistemic state: let $K * * \alpha =$ the result of presenting α to K with $K * * \alpha = K * \alpha$ if α is accepted. In this case, we will obviously have $\alpha \in K * \alpha$. However, if α is not accepted, then $\alpha \notin K * * \alpha$ but

absorption into the belief set. It is an implausible description of the revision process that the moment the new belief enters our belief set, its status undergoes a rapid degradation as far as having any importance in our inference is concerned. There is no notice paid to the fact that it is our most recent piece of information. Often, however, more recent information plays a more significant role in my actions than beliefs formed a long time ago: in the case of the pilot, the belief that he has just observed a friendly aircraft in front of him is more likely to play an immediate role in his actions than his older belief that there are enemy aircraft in the area. What would be desirable would be a model in which some of the importance of temporal ordering amongst beliefs is captured formally[28]: contraction could be plausibly modeled as loss of belief upon receipt of newer information that conflicts with older information and the temporal ordering of information could play a part in inference from beliefs. In the belief sequences model, we will try and address this shortcoming.

- If K consistent, then $K * \alpha = K$ if $\alpha \in K$.

The success axiom does not reflect the relative entrenchment of beliefs: If I already believed α, then receiving new information α often will not leave my belief set as it was before. The new belief set might have the property that it has received additional confirmation of α which it did not possess before. Even epistemic entrenchment assumes the presence of an ordering on beliefs and does not indicate how this ordering might change on receipt of new information. The success axiom is plausible when added confirmation cannot affect the standing of that belief in an agent's belief set.

it still may be the case that $K * * \alpha \neq K$ (this example is due to Rohit Parikh). Hansson has argued that the priority given to new formulas is mistaken and has suggested non-prioritized methods of revision that do not automatically accept new information [51]. Renata Wasserman has suggested a formalization of pre-acceptance states in [104].

[28]It might be argued that the temporal order of beliefs has no bearing on those beliefs that we can term *background beliefs* i.e those beliefs that are likely to play a part in most inferences that we carry out on a daily basis.

One could say that if $\alpha \in K$, then $Cn(K * \alpha) = Cn(K)$ but $K * \alpha \neq K$: doing so makes a clear statement to the effect that an epistemic state is not adequately represented by a belief set. If K is not a belief set, but a richer structure (i.e., with an entrenchment ordering added) then one could still say that $Th(K * \alpha) = Th(K)$, but $K * \alpha \neq K$. In the pilot example, when an enemy aircraft is sighted on the horizon, the pilot receives confirmation of something already believed. This information does not leave his epistemic state unaffected; the pilot now responds and reacts differently by getting his weapons systems ready.

There is a criticism, however, against the success axiom that is misguided. Friedman and Halpern [30] present the example of the scientist who drops two objects of unequal weight from a leaning tower somewhere in Italy and argue that the scientist does not immediately believe in the proposition α that follows: bodies of unequal weights fall at the same rate. Friedman and Halpern take this as pointing to the implausibility of the success axiom. This example, unfortunately, is not very convincing and makes the wrong sort of criticism. Why should the scientist not believe in what has just been observed? Friedman and Halpern confuse the two issues at hand. One is the issue of whether the generalization that follows from the observation should be believed. The observation is: the two bodies fell at the same rate. The generalization is α as above. The answer to the latter is probably in the negative, since more evidence is required to accept the generalization (assuming a reasonable model of scientific investigation). The agent should have no problem with believing the observation of the experiment unless the agent disbelieves its senses. However, if the agent disbelieves its senses, it faces problems far worse than can be solved by any model of belief change.

1.7 The Problem of Iterated Revision

The problem of iterated revision is that of adequately characterizing the process of revising an epistemic state E, obtaining a new epistemic state $E * \alpha$ and then revising again: we want to be able to make comparisons between $(K^*_\alpha)^*_\beta, (K^*_\beta)^*_\alpha, K^*_{\alpha\wedge\beta}$. Matters are not quite so simple: if we have lost the information that we used to construct the set $E * \alpha$, we cannot carry out another revision of the set that results. If the construction tells us how to revise a given belief set but cannot tell us how to describe the intellectual history of a given agent, then it cannot tell us how an agent revises its belief sets in response to one piece of information after another. These problems exist with all explicit constructions for AGM axioms; they make excessively strong assumptions about the information that is present in an agent's epistemic state to guide revisions.

Selection Functions

As a reminder, the representation theorem for selection functions[29] indicates that all work in the revision process is done by the selection function: the revision is described in terms of the selection function. The selection function itself simply takes K, the belief set, as an argument. This means that K is not enough to guide revisions, since the information needed for the revision is contained in the selection function (change in epistemic states is only indirectly dependent on K). The need for a state specific selection mechanism makes iterated revision extremely difficult. Let $E =< K, S >$ be an epistemic state such that K is a belief set and S is the associated selection mechanism. On revision we would like to obtain $E' =< K', S' >$. This requires S to not only determine the change from K to K' but also its own change

[29]See Appendix A.

from S to S'.

Epistemic Entrenchment

In revisions based on epistemic entrenchment (Appendix A) we start with a belief set and an entrenchment ordering $(K, \leq_K)$ and by the end of the revision we are left with a new K', but no entrenchment ordering to go with it (since, in general, $K \neq K^*_\alpha$, the new entrenchment relation $\leq_{K^*_\alpha} \neq \leq_K$). Therefore, this construction does not support iterated revision[30]. To tackle this problem, Rott [92] defines revision of entrenchment orderings as follows:

$$\psi \leq_{K*\phi} \beta \text{ if } \phi \rightarrow \psi \leq \phi \rightarrow \beta$$

Such a definition implies that any further revision of $K*\phi$ will always include ϕ ($\phi \rightarrow \phi$ is a tautology and so is maximally entrenched, and for any $\psi, \phi \vdash \psi \rightarrow \phi$, so that no such ψ can remove ϕ from this position). This means that once we have revised by a particular belief, we can never lose that belief (which seems counterintuitive).

Systems of Spheres

In Grove's systems of spheres construction for AGM revision[31] an $\mathcal{L}$-world is a complete and consistent truth assignment to formulas in $\mathcal{L}$. The set $\mathcal{W}$ is the set of all $\mathcal{L}$-worlds and $\preceq$ is a ranking or total preorder on worlds in $\mathcal{W}$. Now, consider $min_\preceq$, the set of all minimal worlds in $\mathcal{W}$ under this preorder i.e., all worlds w such that there is no $w' \preceq w$. With $min_\preceq$ we associate the set $Bel(min_\preceq)$ which consists of all the formulas that are true in *all* worlds in $min_\preceq$. A revision operator is defined as follows:

---

[30] Some suggested solutions are to use a single entrenchment ordering for all belief sets (as in Rott, [93]), or, to use a function that given a belief set, gives us an entrenchment ordering for the set in question (as in Schlechta, [96]).

[31] As described in Friedman and Halpern, [30]. This treatment is originally due to Boutilier [7] and Katsuno and Mendelzon [58].

DEFINITION 4 [30] *$Bel(min_{\preceq}) * \alpha$ = all formulas that are true in all minimal α-worlds according to $min_{\preceq}$.*

It can be shown that $*$ above satisfies the AGM postulates. Therefore, revision operators are defined by starting with a collection of orderings $\preceq_K$, one for each belief set K under consideration. To define $K * \alpha$ we use the above starting with the ranking $\preceq_K$ corresponding to a particular K (for every set other than the inconsistent set, $Bel(\preceq_K) = K$). Every belief revision operator that satisfies the AGM axioms can be characterized by this method.

Such a method runs into problems with iterated revision since the revision operator so obtained is represented by a family of rankings, one for each belief set. The rankings that represent the epistemic state associated with a belief set give us the information on how to revise, not the set itself. However, since the revision process gives us the revised belief set, not the revised ranking, the representation does not support iterated revision.

Friedman and Halpern suggest a distinction between E (the epistemic state) and $Bel(E)$ (the set of beliefs associated with the epistemic state) and point out that $E*\alpha$ should always be an epistemic state. All constructions considered so far start with an epistemic state (a belief set plus ranking, or selection function, or entrenchment ordering) and, after revision, give us a set of formulas (i.e., the revision functions are mappings from epistemic states and formulas to belief sets). Furthermore, to have successful revisions, it should be the case that $\alpha \in Bel(E * \alpha)$ and in cases when expansion can be meaningfully defined[32], it should be the case that $Bel(E * \alpha) \subseteq Cn(Bel(E) \cup \{\alpha\})$. This would require that revision always be contained in the

[32] We mention this caveat since in the direct approach to revision, there is no clear counterpart to the expansion operation.

simple expansion of the epistemic state.

Iterated Revision Proposals

Boutilier [6] proposes the 'natural' revision operator: revision by a *sequence* of formulas is equivalent to revision by the conjunction of the formulas. For a belief set K, revising by $\alpha_1, \alpha_2 \ldots \alpha_n$ is the same as revising by $\alpha_1 \wedge \alpha_2 \wedge \ldots \alpha_n$ ($K * \alpha_1 * \alpha_2 \ldots * \alpha_n = K * (\alpha_1 \wedge \alpha_2 \wedge \ldots \alpha_n)$). If we revise by a formula that is inconsistent with this sequence, we drop all formulas starting with the first one that it conflicts with. This is a rather severe form of revision and seems susceptible to counterexamples that stress its implausibility. In our pilot example, according to the natural revision operator, the pilot would stop believing that the aircraft observed was a two-seater upon noticing that it was not a hostile aircraft. The sequence of the observations would make a difference since had the pilot observed that it was a two-seater aircraft before classifying it as hostile, he would have continued to believe it was a two-seater after realizing it was a friendly craft.

As in the case of the trivial update, there is an absence of any notion of relevance sensitivity in the natural revision operator. The belief that the aircraft is a friendly one is of no relevance to the belief that the aircraft is a two-seater, and yet the latter would be dropped on acceptance of the former. It might be relevant if the pilot knew that his air force only flew single seaters, but in the absence of any such background information, it seems perverse to drop the older belief[33].

Darwiche and Pearl

Darwiche and Pearl [16] put forward a set of postulates that deal explicitly with

[33] In the B-structures model, we will show that the pilot will only revise beliefs about the nationality of the aircraft and not its seating capacity, and, in the case of sequences model, it will be shown that the pilots older belief that the aircraft is a hostile craft will be dropped (while beliefs about its seating capacity remain untouched) because of the receipt of newer, relevant information that the aircraft is not hostile.

iterated revision ($\circ$ is the the update operator):

(C1) If $\alpha \models \mu$, then $(\psi \circ \mu) \circ \alpha \leftrightarrow \psi \circ \alpha$.

(C2) If $\alpha \models \neg\mu$, then $(\psi \circ \mu) \circ \alpha \leftrightarrow \psi \circ \alpha$.

(C3) If $\psi \circ \alpha \models \mu$, then $(\psi \circ \mu) \circ \alpha \models \mu$.

(C4) If $\psi \circ \alpha \not\models \neg\mu$, then $(\psi \circ \mu) \circ \alpha \models \mu$.

The postulate **C1** states that when two pieces of information (one more specific than the other) arrive. the first is made redundant by the second. **C2** says that when two contradictory epistemic inputs arrive. the second one prevails: the second evidence alone yields the same belief state. **C3** says that a piece of evidence. μ should be retained after accommodating a more recent evidence α that entails μ given the current belief state. **C4** simply says that no epistemic input can act as its own defeater. However. Arlo-Costa and Parikh [4] and Lehmann [63] have shown that **C2** is problematic in allowing *only the trivial update.* This is done by considering the case when the formulas in question are not simple propositional atoms. Freund and Lehmann [29] show that it is inconsistent with the AGM axioms[34].

The example above. problems with the iterated revision proposals [16. 63. 6. 29] and the AGM trivial update point us in the direction of trying to incorporate some notion of relevance into revision operations that ensure minimal change and informational inertia. A more vigorous point is made by Goldszmit and Pearl [42] who argue that the best reason to reject the entire AGM framework is its failure to handle iterated belief revision. While not rejecting the AGM framework. we will take the approach that the best way to handle the problem of iterated revision is to consider the direct method of belief revision.

[34] As is the weaker axiom, $C2'$ proposed in Nayak *et. al.* [72].

1.8 Alternatives: Belief Sets versus Bases

The stage is now set to contest the AGM model on a fundamental issue: the insistence on the logical closure of the belief set associated with an agent's epistemic state. This issue is raised by proponents of the *belief base* method: Fuhrmann [31], Nebel [74, 75, 76], Makinson [69], Hansson [47, 48]. The central objection is that the coherentist AGM model ignores the distinction between explicit and implicit beliefs. There are beliefs which are explicit - they may be actually asserted or at least agreed to. Other beliefs are implicit in the sense that an agent can be made to agree to them after a discussion. Beliefs can be implicit too, in that they might be inferred with minimal effort from an agent's beliefs or they might be implicit in the fact of having an explicit belief in another fact. It seems plausible then, that an agent's beliefs are not logically closed[35] and that there is an interesting distinction to be drawn between what the agent explicitly believes and what can be derived from that set. This distinction can be made with respect to implicit beliefs as well, in that there is an interesting distinction between beliefs that are implicit in being formally derivable from the set of explicit beliefs and those that do not require such a derivation. Working with belief sets, with its lack of demarcations amongst beliefs, ignores the fact that some beliefs function as *reasons for other beliefs*. To maintain this distinction is to adopt an approach approximating the foundationalist approach as mentioned above (though with not the fervor that the foundationalist has in its basic set of beliefs).

In the coherentist approach, if the set of explicit beliefs is consistent, then we

[35]It is worth pointing out the suggestion made by Fred Dretske and Robert Nozick that the *principle of epistemic closure for knowledge* be dropped. The motivation for a suggestion is a resolution of the skeptical puzzle, but there are similarities with our current discussion in that adopting the principle of epistemic closure commits us to describing the agent has possessing more knowledge than the agent can realistically possess.

can define the set of implicit beliefs as just their logical consequences. However. this becomes implausible[36] if, known or unknown to the agent, the explicit beliefs are *inconsistent*. The relaxation of the condition of logical closure on states represented by bases provides a significant feature that is of special relevance to this project: *bases let us tolerate inconsistencies while making a distinction between the different beliefs of two agents.* Two different bases can give us the same belief set while being different in content. Consider the following two bases:

$$H_1 = \{\alpha_1. \alpha_2 \rightarrow \beta_1. \alpha_2 \wedge \beta_3. \neg\beta_1. \alpha_2\}$$

$$H_2 = \{\alpha_1. \alpha_2 \rightarrow \beta_1. \alpha_2 \wedge \beta_3. \neg\beta_1. \alpha_2. \beta_1 \wedge \beta_2 \wedge \beta_3\}$$

According to the closure conditions on belief sets. both these agents have identical beliefs since $Cn(H_1) = Cn(H_2) = K_\perp$. However, this does not seem to be the case. The two agents have palpably different beliefs in that we can point to a proposition that is not contained in one agent's belief base while being contained in the other agent's belief base. Consider the case of my brother and me. We both are often inconsistent, especially with respect to our political opinions. Yet, there is no way that we could be described as having the same beliefs as we would be if our epistemic state was to be described by closure conditions: my brother is inconsistent with respect to his views on criminal justice and I am inconsistent with respect to my views on socioeconomic policy. In general, it is possible for two bases $H_1. H_2$ to differ, have $Cn(H_1) = Cn(H_2)$, and yet have $H_1 * \alpha \neq H_2 * \alpha$: *syntax* is of prime importance in belief bases.

It is in the proposals for belief bases that we find the first proposals for allowing

[36]Such a definition is also implausible for reasons having nothing to do with inconsistency: as our discussions above indicate, identifications of implicit beliefs with the consequences of explicit beliefs do not capture all our intuitions about the distinctions between the two.

inconsistencies in belief representations. Fuhrmann [31] takes seriously the possibility of allowing inconsistencies in belief states and argues that two desiderata should be satisfied by any realistic model of belief revision:

- Localize inconsistency: an inconsistent theory should not be totally corrupted by virtue of the infection of the entire theory.
- Locally restore consistency: resolve one inconsistency at a time by contracting an inconsistent theory so that other inconsistencies which cannot be presently resolved carry over into the contracted theory (repeat this process as required).

Base Contraction and Belief Set Contraction

Hansson has argued that modeling epistemic states as bases leads to a more intuitive set of contraction operations.

Example: I believe that *The Bicycle Thief* is playing at the Anthology Film Archives tonight (α). I conclude that either *The Bicycle Thief* or *Shoeshine* is playing at the Archives tonight ($\alpha \vee \beta$). If I then find out that *The Bicycle Thief* is not playing, then I also lose my belief in $\alpha \vee \beta$ since this is no longer implied by the belief base (which was originally α).

In belief set contraction however, this would have to have been ensured by use of a selection function. In base contraction, the selection function is left with the work of conflict resolution amongst basic beliefs and not the derived beliefs i.e., it is relieved of most of its work. Hansson [47] provides another example (modified below for conciseness) to demonstrate the distinction that bases let us make amongst belief sets which correspond to distinct belief bases i.e., $K = Cn(H_1 = \{\alpha\}) = Cn(H_2 = \{\alpha, \alpha \vee \beta\})$[37].

[37] We can also think of this example as illustrating the contention that using bases as representa-

Example: I live in a town that has two movie houses. When I meet a friend who holds a ticket stub in her hand, I draw the conclusion that at least one of the movie houses is open: $\{\alpha \vee \beta\}$. From a distance I see that the marquee lights for one house are on. I believe that that movie house is open today, $\{\alpha\}$. This belief state can be represented by $\{\alpha, \alpha \vee \beta\}$. When I reach the movie house, however, the movie house is closed (the marquee lights were just turned on for advertisement). Now, I believe β. If I had not met my friend, though, my only clue would have been the lights. The original belief state would have been $\{\alpha\}$ and after finding out that the movie house was closed, the resulting set should not contain $\alpha \vee \beta$ since in that case there would have been no reason to believe that one of the movie houses was open[38].

Approaches to Revision on Bases

We can think of bases in different ways. On one reading the epistemic state is a belief base itself and all revision operations operate directly on the base itself (i.e., there is no closure of the epistemic state). The particular syntactic representation of the base then guides the revision. Revision operations defined on such entities are called *base revision operations.* The result of any such operation should be another belief base (to preserve categorial matching) and postulates for these operations can be provided (see Hansson [47]). Base revision operations can be matched to corresponding belief set revision operations by considering the belief revision operation that corresponds to an operation on the base that generates the belief set. These are called *base generated belief revisions.*

On the second reading, the epistemic states are still belief sets, but finitely axiomatized by a given belief base. So, for a set K, we are given a base H such that

tions for epistemic states provides a notion of *local change*; belief sets are too tightly interconnected.

[38]This example assumes that agents retain reasons for their beliefs over time.

$K = Cn(H)$. The changes on the belief set are guided by the content of the belief base. This second approach suffers from the problem of iterated revision in that the result of a revision operation on a base H may be a belief set that might not be axiomatizable as a base. So, while we revise we have a base, but after revision, we might not have a base to work with for the next revision.

Revisions on bases are especially sensitive to the syntax of the representation: two bases that have the same consequence sets can have different states after revision. When we want the formulae in the belief state to reflect distinctions between explicit and implicit beliefs, and when we want to be able to make meaningful distinctions between states which would be equivalent under logical closure, the syntactic approach makes intuitive sense.

1.9 The Computational Complexity of Belief Revision

How computationally realistic are belief revision schemes? One argument against using belief sets as representations of belief states is their size. A belief set over a finite language can have a size that is double exponential in the size of the language. Furthermore, the logical closures of sets of beliefs are infinite and so, are representationally infeasible: an issue of key importance in constructing a plausible model of reasoning or implementing a belief revision system[39].

Some precision is desirable in order to discuss the *computational complexity* of

[39] Another problem (as pointed out by Nebel [77]) is that whenever a constructive method (such as epistemic entrenchment or meet contraction or the spheres method) uses extra-logical preference information, it assumes a relation that extends over the set of *all formulae* in a logically closed theory, or a relation that extends over *all models* of a theory, or a relation that extends over *all subsets* of a theory.

belief revision[40]. Take n to be the size of a problem instance and let f be a monotonically increasing function ($f : Z^+ \to Z^+$). The computation time of the problem is bounded by f if for all instances of the problem, no more than $f(n)$ computations are necessary. The computation time of the problem is *polynomially bounded* if there is a k such that the computation time of the problem is bounded[41] by $O(n^k)$, and is *exponentially bounded*[42] if there is a $b \in Z^+$ such that the computation time of the problem is bounded by $O(b^n)$.

The problem of determining the membership of a sentence ψ in a belief set $K = Cn(C)$ revised by ϕ is the *derivability* problem:

$$\psi \in K \dot{+} \phi$$

The input size of the problem is a function of the size $|C|$ of the belief base, and the sizes $|\phi|, |\psi|$ of ϕ, ψ. To calculate the complexity of this problem we calculate the number of operations required to determine whether ψ is derivable from the set $K \dot{+} \phi$. In the propositional case, we test whether the set $K \dot{+} \phi \cup \{\psi\}$ is satisfiable. To check for satisfiability of a set of propositions consisting of n distinct propositional symbols is to check 2^n truth assignments: the solution requires exponential time.

This means that this most straightforward problem in propositional logic is one that is *computationally intractable*: a problem is considered *computationally tractable*

[40]This discussion draws on Nebel's survey of the complexity of belief revision in [77].

[41]If f and g are functions, c is a constant, then $O(f)$ is defined to be the class of functions g such that:

$$\forall n, g(n) \leq c \times f(n)$$

[42]The difference between polynomial and exponential solutions for a problem is significant. Consider a problem solved by an algorithm that is of the order $O(n)$ and also by one that is of the order $O(2^n)$ on a machine whose speed is a 1000 operations per second. The first algorithm can accept an input size of 1000 in one second, while the second is restricted to an input size of $\simeq 9$. Increasing processing speed by a factor of 1000 increases the input size for the first algorithm by 1000 but only by $\simeq$ 9-10 for the second algorithm.

only if for all instances of the problem, an algorithm can solve the problem in a number of steps that is polynomially bounded. This assumes a sequential, deterministic model of computation i.e., the classical Turing machine. This class is denoted **P**. The class of decision problems that are accepted by nondeterministic Turing machines using polynomial time is called **NP**. Since the class of deterministic machines is contained in the class of non-deterministic machines, it follows that $\mathbf{P} \subseteq \mathbf{NP}$. A problem that has the property that all problems in **NP** can be polynomially reduced[43] to it is at least as hard as all problems in **NP** i.e., it is **NP**-hard. A problem that is in **NP** and is **NP**-hard is **NP**-complete (if a polynomial algorithm exists for any problem in this class, it exists for all of them). The classical satisfiability problem is an **NP**-complete problem. Its complement **UNSAT** as well as propositional implication (the basic derivability problem i.e., whether $\psi \in Cn(\emptyset) \dot{+} \phi$ is co-**NP** and tautology checking are co**NP**-complete.

The complexity of belief revision procedures can be placed in a hierarchy of complexity classes which classify problems harder than **NP** and lie inside a class of problems (**PSPACE**) solvable in polynomial space on a deterministic Turing machine. Between **NP** and **PSPACE** there (probably) exists an infinite hierarchy of complexity classes (problems of the same complexity lie in the same class) called the polynomial hierarchy (PH). Nebel [75] shows that belief revision problems fall into the lower end of the polynomial hierarchy i.e., the best algorithms are all at least as hard as the satisfiability problem and somewhat harder.

Let X be a class of decision problems; $\mathbf{NP^X}$ denotes the class of decision prob-

[43] A problem A can be polynomially reduced to another, B, (written $A \leq_p B$) if there exists a function f that can be computed in polynomial time and has the property that $x \in A$ if and only $f(x) \in B$. That is, an algorithm for problem B can be used for A with only minimal (polynomial) overhead. So, B is at least as hard as A with respect to solvability in polynomial time.

lems $L \in \mathbf{NP}^{\mathbf{X}}$ such that there is a nondeterministic Turing machine that solves all instances of L in polynomial time using an *oracle*[44] for $L' \in X$. We define the classes $\Delta_p^k, \Sigma_p^k, \Pi_p^k$ as follows:

$$\Delta_0^p = \Sigma_0^p = \Pi_0^p = P;$$
$$\Delta_{k+1}^p = P^{\Sigma_k^p};$$
$$\Sigma_{k+1}^p = NP^{\Sigma_k^p};$$
$$\Pi_{k+1}^p = co - \Sigma_{k+1}^p$$

Thus, $\Sigma_1^p = \mathbf{NP}$, $\Pi_1^p = \mathbf{coNP}$. Nebel then obtains the following results for the trivial update and model-based revision schemes ([14][45]):

THEOREM 1.3 [75] *The trivial update is in* $\Delta_2^p - (\Sigma_1^p \cup \Pi_1^p)$, *provided* $(\Sigma_1^p \neq \Pi_1^p)$.

THEOREM 1.4 [75] **MBR** $\in \Delta_2^p - (\Sigma_1^p \cup \Pi_1^p)$, *provided* $(\Sigma_1^p \neq \Pi_1^p)$.

Nebel demonstrates that those revision procedures that do not satisfy all the revision postulates, cannot be shown to be in this class[46] and concludes that the best result for a belief revision problem is membership in Δ_p^2.

Potential Solutions

Two approaches can be used to solve the problem of representational feasibility. One is to assume that there is no preference information: as in the trivial update and model based revision schemes. Yet another is to attach preference information to the formulae in a belief base by providing a preference relation over the formula or by partitioning the base into *priority classes*. Such methods are called syntax-based revision schemes. In either case, the following results obtain:

[44]An oracle enables us to decide any problem in a class of decision problems in unit time i.e., it provides answers without cost.
[45]For a survey of the complexity of model based revision schemes, see Eiter and Gottlob [20].
[46]Nebel, [75].

PROPOSITION 1 [77] *Any representationally feasible revision scheme is* **NP**-*hard and* **coNP**-*hard. and it is not a member of* **NP** $\cup$ **coNP** *(provided* **NP** $\neq$ **coNP***).*

PROPOSITION 2 [77] *Any representationally feasible revision scheme is* **NP**-*hard and* **coNP**-*hard even if the size of the revision formula is bounded by a constant.*

Nebel obtains the following result for prioritized[47] base revision:

THEOREM 1.5 [75] *The problem of deciding whether* $H * \alpha \vdash \beta$ *where* $*$ *is a prioritized base revision. is* Π_2^p*-complete.*

If a problem is completeness for the class Π_2^p then the problem is solvable in polynomial time if and only if **NP** = **P** and the best known algorithms are exponential in the worst case. Using a more tractable background logic such as Horn logic (i.e.. all propositions are Horn clauses[48]) is not enough to get an instance that can be solved in polynomial time. Another solution is to restrict the cardinality of priority classes to a fixed value (for example. to singletons). In that case. the general problem becomes **NP**-equivalent and polynomial for Horn clauses. Eiter and Gottlob [20] suggest placing a restriction on the size of the update formula. This does not affect membership in the class Π_2^p either.

To sum up, most belief revision procedures are hopelessly intractable. Nebel suggests that there are two possible ways out. One is to tolerate incompleteness in the semantics of the logic i.e., to use a multivalued logic which makes only logically valid conclusions but misses some of them. The other is to reduce the size of the problem instance. Our approach to the problem of feasible belief revision will incorporate both of these suggestions.

[47]See Appendix A.
[48]Disjunctions in which there is at most one non-negated literal.

In conclusion, problems with the AGM method can classified as follows:

- Problems with the plausibility of the rationality postulates and their counter-intuitive consequences.
- Problems with constructions that block iterated revision.
- Problems with the adoption of the coherentist approach.
- Problems with the representational infeasibility and computational intractability.

These problems are not of equal severity for the enterprise. Problems with the postulates can be addressed within the AGM framework itself. The problem of iterated revision, however, will require us to adopt the direct method of belief revision. Perhaps more fundamental a problem is the adoption of the notions of consistency and logical closure of the belief state. The wholesale adoption of these constraints have led us to consider the dropping of logical closure in the case of belief bases. In the next chapter I continue a line of argument begun in this chapter to establish the following theses:

> **I.** A realistic model for belief representation and belief revision must pay primary attention to the notions of minimal change, inconsistency tolerance, computational tractability and the distinction between explicit and implicit beliefs;
>
> **II.** The first three notions are all dependent upon some notion of relevance being present in the representation and revision procedures.

Chapter 2

Inconsistency Tolerance

2.1 Rejecting Classical Models of Reasoning

The central problems with the logic-constrained approach are its insistence on the consistency and closure of an agent's epistemic state and hence, its close alliance with classical notions of rationality. Such an alliance is understandable in that belief revision theorists often take themselves to be offering a normative theory of the logic of belief change; such a theory is guided by intuitions of the form that 'rational' belief change should take. However, for the purposes of a *realistic* theory of belief change, one must look elsewhere. We begin by rejecting classical models of reasoning and belief change in favor of ones that pay attention to the *cognitive constraints* on human agents and enable finer distinctions amongst epistemic states. To do so is to argue for the following theses:

I. An agent's beliefs are not closed under deduction.

II. The inconsistent belief state is not vacuous or irrational.

I, II directly concern themselves with the proscription of inconsistency of classical belief states. The combination of the two above will entail that the presence of a contradiction does not entail the vacuous belief set[1]. To argue for these theses is to argue for a paraconsistent logic as well (without committing ourselves to saying that a paraconsistent logic is the best way to model human reasoning) and to argue for a particular model of how we take our beliefs to be structured. The primary argument is that for inconsistency: we have already provided preliminary arguments against the logical closure of belief states in the previous chapter.

Should we also argue for the following?:

III. No human beings are rational as described by the classical model.

III however, argues for a different position and one with no effect on the normative theorists position: the mere nonexistence of agents that reason in accordance with the classical model does no damage to its normative import. Therefore, the position suggested by **III** is not one that we will adopt in this study. The classical model of rational reasoning can still be taken as a standard, while we concern ourselves with a reasonable description of how actual agents reason. The point of interest for us then becomes the amount of deviation from this normative standard.

To argue for the theses above is to argue that inconsistent beliefs could be entertained by a rational reasoning agent: to argue that since inconsistent beliefs are pervasive in all human reasoning, *a plausible model of belief representation and revision must be one that allows for the entertainment of inconsistent beliefs.* The logic of inference that this model will suggest for the agent in question should cause us to investigate the plausibility of traditional sound rules of inference such as adjunction, keeping in the mind that it is the supposed soundness of these rules that makes us impose the standards of logic on real world reasoning. If our models can help us

[1]If we use classical logic to model the reasoning of an inconsistent agent, then of course, the agent will possess the vacuous belief set, but the point of this argument is that we need a different logic for inference from beliefs. Logic for epistemology, not classical inference is our concern.

in the resolution of classical puzzles about belief then so much the better for their plausibility.

2.2 Implications for the Modeling Enterprise

What implication will the arguments above have for the enterprise of constructing models of human reasoning? The answer lies in how we view our tasks. There are two sorts of ways to conduct this entire endeavor:

Model Primacy: One can define a prescriptive model of reasoning antecedent to a determination of actual human behavior. If human reasoning is not described by the model, we can describe human agents as irrational. The fault or the shortcoming (if it can be described as such) lies *not in the model itself but in human behavior.*

Empirical Primacy: One can define a model of reasoning, determine actual human behavior and if human behavior or reasoning is not described by the model, scrap the model and come up with one that does justice to the 'data' on the way human beings reason. Such a method is iterative, seeking feedback from the 'data' to guide it in its construction of models.

The strategy that we choose out of the two above will largely be determined by our take on the enterprise of cognitive evaluation and modeling. Is our model of reasoning intended to be purely descriptive i.e., should only the empirical guide us in the construction of our model of rationality? Is it intended to be purely prescriptive i.e., should theoretical intuitions be our only guide in model building? If our response is the former of these two, then we must take the empirical to be primary. If our response is the latter, we maintain the sanctity of the model, and resign ourselves to describing human behavior as irrational if it does not conform to the strictures of the model. I suggest that either of the two extremes, i.e., a purely prescriptive model or a purely descriptive model is implausible. Existing purely prescriptive models have very poor fit with psychological facts about (and limitations on) human cognitive performance, and purely descriptive models do not pay sufficient attention to intuitions about what constitutes rational cognitive performance.

Inspiration from the AI Logicality vs. Rationality Debate

A putative clash between the prescriptive and descriptive models occurs in the disputes between work in artificial intelligence and purely normative models of reasoning. Here, 'rationality' as defined by artificial intelligence researchers is not the rationality that philosophers engaged in the task of providing normative models of reasoning normally work with. As an illustration, consider the following definition of *rational inference* by Doyle [18]:

> [A]n approach to reasoning from inconsistent knowledge. In this approach, the agent rationally chooses a consistent subset of its explicit beliefs, and then uses this subset to choose a consistent set of implicit beliefs.

In AI, the main theories for reasoning with inconsistent knowledge aim to capture a certain amount of psychological fit with a plausible model of human reasoning rather than aiming for a normative ideal. Which idea brings us to the notion of the psychological plausibility of a logic: the degree to which a logic of reasoning captures the reasoning patterns that people ordinarily use. Are the rules of inference of the logic ones that rational agents would employ? Our response is that if a plausible method of belief representation and revision suggests a logic that loses classical rules of inference, it should serve as a point of examination of the applicability of classical rules of inference to epistemic contexts. If we lose rules of inference required for the logical closure of belief sets in our logic, then the loss of these rules further bolster the arguments against the closure of belief sets. If there are rules at work that ensure the presence of the absurd belief set when an inconsistency is present, then the loss of those rules, as a way of ensuring the non-absurdity of an inconsistent belief set is a good thing. Since agents do not need all the logical consequences of their beliefs in order to function coherently, the logic of inference for epistemic contexts is quite different from classical inference. We take certain rules of classical inference to be sound since they aid us in making plausible inferences: it is merely the unrestricted application of those same rules that I take to be problematic. The solution to this clash between classical rules of inference and the rules suggested by an alternative model of inference is *restricted application* of rules of inference, dependent on something other than just pure logical form and dependent on the subject matter of the propositions involved.

2.3 Making the Case for Inconsistency

Despite the existence of the consistency and closure conditions, a variety of sources in the philosophical and psychological literature (see [78, 100, 103, 84]) have presented empirical evidence and philosophical arguments to show that humans do entertain inconsistent beliefs and these beliefs are not closed under deduction. These agents are not people that we take to be irrational; if they were, we could find ourselves confronted with a situation where ordinary agents, non-pathological in every sense of the word, would be described as irrational by classical models of reasoning.

The requirement of checking for consistency with old beliefs on receipt of new information is not intended as a psychological description: a reasoner does not solve a computationally intractable task each time it adds a new fact to its belief set. Since we do have some introspective idea of whether we are consistent or not, this suggests that we do checks for consistency, but only in some limited sense. What role does the condition for consistency of the belief state play?

We can think of an agent's consistency in terms of what Levi [65] terms a 'commitment' to consistency: there is a tendency on the part of the agent to contract its epistemic state by those beliefs that cause inconsistency and to expand its epistemic state by those that would make it consistent. The agent is perhaps never done with its task and it is this agent that we take ourselves to be describing. There is a difference, then, between what the agent's doxastic commitments are and what its doxastic *performance* is. Levi, who makes this distinction, suggests that an agent *cannot* satisfy the rationality demand when it comes to doxastic peformance[2]. We can take ourselves to be respecting the constraints of the doxastic commitments of the agent while simultaneously describing its doxastic performance in the models of reasoning that we construct for it.

Examples of Inconsistency

Inconsistent beliefs appear to be ubiquitous in human reasoning: limitations of memory and reasoning capacity often make this the case. There are other reasons as well. For example, I think that quantum mechanics (**QM**) is a true theory: I also think that General Theory of Relativity (**GTR**) is a true theory. I also know that the two are in deep and fundamental conflict with each other and cannot be compatible with each other without significant modifications. Yet, I refuse to abdicate belief in either of them[3]. The inconsistency simply seems too far away from empirical

[2]This is the *cognitive economy* argument against logical closure and consistency: the expenditure of cognitive resources required for such a facility is prohibitive (Rescher [89], Harman [53], Levi [66], and Cherniak [9]).

[3]It might be said in objection that scientists do not really believe in a theory, they simply accept

testability to be of any relevance to the rest of my beliefs. Similarly, inconsistencies in one subset of our beliefs do not infect the entire set of our beliefs with absurdity. In other cases, agents could unknowingly hold false, inconsistent beliefs and give one of them up on being informed of their inconsistency. This revision does not constitute an argument against an agent holding inconsistent beliefs consciously. For even if we were to realize that our beliefs were inconsistent, we might not revise our beliefs immediately. Revising our beliefs in a fashion that renders the remainder of the belief set consistent is a non-trivial task and one that requires the expenditure of time. While we consider strategies for belief revision, we are stuck with inconsistent beliefs (such as in my deciding to hold on to my beliefs in both **GTR** and **QM** till the unified theory comes along).

Our example above provides a simple argument against the demand for consistency: to demand consistency in certain fields of knowledge would be to actually *suspend* belief in any of those fields since they are so clearly infected with inconsistency. Perhaps inconsistency in beliefs is a pervasive state of affairs, a desirable attitude to have in the face of often conflicting information that we receive about the world. Perhaps the entertainment of inconsistency is a reasonable and often valuable cognitive strategy. There are other, well-known, examples of inconsistency in the philosophical literature (these also serve as arguments against the deductive closure of a rational set of beliefs).

The Paradoxes of Inconsistent Beliefs

The Lottery, Preface and Sorites paradoxes are the best known illustrations of the pervasiveness of inconsistent beliefs in our thinking: each of the examples reveals the

one to work with. However 'accepting to work with' and 'believing' are not so easily distinguished especially in terms of the scientist's commitment to certain actions and assertions.

benign nature of the inconsistency they contain.

The Sorites Paradox

In the Sorites Paradox if i is **few**, then $i+1$ is **few**...iterating i n times (for some n which is a large number) we arrive at the conclusion that i_n is **few** where i_n is clearly a large number. The conjunction of these premises is clearly inconsistent. Yet, this inconsistency is not one that renders us irrational if we are to tolerate it[4]. Vagueness in ordinary language is something that we live and work with, recognizing that unrestricted application of its predicates will involve us in inconsistencies that are harmless. I could respond to a set of queries about my color judgements that would show me to be inconsistent, and yet, not consider myself irrational. My color judgements are, quite naturally, irrelevant to the question of whether I believe in the fairness of a progressive income tax.

The Lottery Paradox

In the Lottery Paradox, originally proposed by Kyburg [60], there is a very high probability ρ that a particular lottery ticket t will not be the winning lottery ticket. So the following hypothesis is acceptable (for any ticket t): 'ticket t will not win'. The conjunction of n such statements for tickets $t_1,...,t_n$ is clearly false: the lottery is fair and therefore the following is acceptable: 'some ticket t will win'. The Lottery paradox suggests that what creates the inconsistency is the unrestricted application of the rule of adjunction: a more plausible logic would be one that would block such an unrestricted application.

The Preface Paradox

In the Preface Paradox, there is a series of statements $S_1, S_2, \ldots S_n$ in a book.

[4]The problem of vagueness in ordinary language leads to demands for multi-valued logics: see Williamson, [108] who concludes that the nature of vagueness is not captured by any approach that generalizes truth functionality.

each of which is individually asserted. The preface modestly asserts that not all the statements in the book are true: $\neg(S_1 \wedge S_2 \ldots \wedge S_n)$. The resulting overall set $(S_1, S_2, \ldots S_n, \neg(S_1 \wedge S_2, \ldots \wedge S_n))$ is inconsistent. Yet, it is plausible to believe the whole set and this is not judged as being irrational.

In the cases above, the descriptive statements are inconsistent and yet rational agents find these descriptions plausible and entertain the inconsistency without losing the ability to function coherently. In the case of the Lottery and Preface paradoxes, the unrestricted application of the rule of adjunction to a large number of statements gets us into trouble. Since a belief set contains statements that have not always attained the status of knowledge, it may be argued that logical closure for those statements be dropped[5].

Against Vacuous Belief Sets: Relevance Relations

The simplest argument against a conception of rationality that commits human agents to perfect consistency is that holding inconsistent beliefs does not render our belief set trivial. I might believe that I am a human being *and* an alien and yet not have the vacuous belief set. I don't believe for instance, that gold is a gas. Why is this not the case? The answer suggests itself in the way we try and explain our rationale. We do not consider the belief that I am a human being *and* an alien to be *relevant* to the belief that gold is a gas.

[5]Inconsistency of a slightly different kind may be found in mathematical theories as well. In naive set theory, the following axioms (intuitively plausible) can be formulated: the *Comprehension Schema* and the *Extensionality Principle*:

- $(\exists y)(x)(x \in y \leftrightarrow A)$
- $(x)(x \in y \leftrightarrow x \in z) \rightarrow y = z$

These axioms, however, entail an inconsistent theory (A is an arbitrary sentence above). From the Comprehension schema, we obtain for some object $r : (y)(y \in r \leftrightarrow y \notin y)$ $(A = y \notin y)$. Now we also get $r \in r \leftrightarrow r \notin r$. It follows that $r \in r \wedge r \notin r$. The solution of these problems by the Restricted Comprehension Schema is often felt to be ad-hoc and a solution born of mere convenience.

This tells us why it is that an inconsistency in our beliefs does not render our belief set absurd: we often do not take beliefs on one subject matter to be relevant to beliefs in another subject matter. Classical models of rationality seem to suggest *trivial relevance* however: every belief is relevant to every other. Introduce an inconsistency, and the entire set is corrupted. The only reason to think that belief in a contradiction entails (epistemically) a vacuous belief set is if one were to believe that all propositions believed were relevant to one another. The moment we reject that contention, we are free to reject the notion of the vacuity of an inconsistent belief state. Such an approach is taken by some paraconsistent logics such as relevance logics. I provide a brief description of the Anderson and Belnap system [3] of relevance logic in the next chapter.

Does Paraconsistency Help?

If $\vdash$ is a relation of logical entailment, it is said to be *explosive* if for all α, β the following relation holds:

ex falso quodlibet: $\alpha, \neg\alpha \vdash \beta$

Paraconsistent logics are not explosive however[6]. This takes care of the objection that an inconsistent belief set would be vacuous. In such a logic the inference above would be blocked and a set of sentences could be inconsistent without entailing all sentences.

If our logic for epistemic contexts was paraconsistent, though, why would we ever *revise* our beliefs? When new information presents itself for our consideration, we could just add it without any internal checks for coherence with the rest of our

[6] Relevance logics [3], and non-adjunctive systems [89, 98] are examples; see Priest and Routley [84] for an introduction and survey.

beliefs. This argument, however, presumes that logical consistency is the *only* factor guiding belief change:

> Nothing forces us to delete $\neg\alpha$ from our beliefs if we receive α, and yet we might choose to delete $\neg\alpha$. Why would this ever be the case? It would be the case when entertaining $\neg\alpha$ would not be *rationally possible* as opposed to just being *logically possible*[7].

Why might we choose to revise our beliefs when it is logically possible to accept all beliefs? Why would we ever think that we might need a logic that tolerates inconsistencies? One answer is that our processes of belief formation and retention depend upon multiple criteria. A belief might be deducible from something already accepted, it might have experimental support, it might have high statistical probability, it might be the only information we have to guide our actions and so on. We might have a pair of inconsistent beliefs that score highly on these counts. We might have a pair of inconsistent beliefs that score highly on methodological grounds: simplicity, problem solving capacity and so on. Their rivals might have poorer problem solving capacity and have faced less tests.

An inconsistent set of beliefs can be rationally acceptable; it can have high value for us by virtue of these other criteria. Of course, we could have an inconsistent belief set that is not rationally acceptable. Its inconsistency might cause problems with other inferences we would like to draw and it might be a poor candidate when evaluated on extra-logical criteria. Revising at all times by the new information that is presented to us can often put us in such a situation: we could lose the simplicity, economy and unity that we would like in our beliefs. We might find this to be an irrational strategy despite the fact that it is logically consistent[8]. Priest suggests that consistency may

[7]Priest, [83].
[8]Ibid., page 7.

be viewed as merely one of a list of potentially conflicting desiderata. Some additions to our belief set render it inconsistent, but at a tolerable cost. It certainly costs me nothing to believe both **QM** and **GTR** while acknowledging that the two are in conflict with one another. The desire to maintain consistency is present, but does not take precedence over several other factors. If that is the case, and we have chosen to tolerate inconsistency in our belief set for non-logical reasons, why think that the simple addition of any and all beliefs is the most rational course of action?

It might be rational to maintain an inconsistent set since our cognitive objectives are often varied and not just restricted to the maintenance of consistency:

> The keystone of cognitive rationality is the idea of doing as well as we possibly can in the cognitive enterprise–of optimizing our attainment of its defining objectives. The maintenance of consistency–desirable though it may be–must be subordinated to this ruling *telos.*[9]

The nemesis of an inconsistent belief set is not just its inconsistency, but its lacking other features that make belief sets desirable. One might suggest that possessing consistent beliefs is neither necessary nor sufficient for a belief set to be deemed desirable. There are too many other factors at play to give mere consistency such a clinching role.

I now consider the strongest suggestion for how inconsistent beliefs *could be entertained* in a plausible model of reasoning: via *compartmentalization.* Beliefs are not presented to us as a monolithic set of statements or propositions. Instead beliefs are are partitioned into sets of beliefs, different subsets of which are activated on different occasions (Cherniak [9]).

[9] Rescher and Brandom, [89], page 44.

2.4 Cherniak on Minimal Rationality

Cherniak's model of minimal rationality is intended to do justice to our limitations as human beings. A general *ideal* rationality condition is as follows:

Ideal Rationality Condition: If A has a particular belief-desire set, A would undertake *all* and only actions that are appropriate[10].

This is unacceptably stringent since it excludes the possibility of someone being confused, mistaken, indecisive, suboptimal in their choice of actions. It entails that the agent has an *ideal capacity* for inference: if the agent had a particular belief-desire set, then the agent would make all and only sound inferences from the belief set that are appropriate[11]. This ability is possessed by those reasoning agents that have the logical ability to select all and only those inferences to make from their beliefs that are appropriate for the agent to make and then successfully perform. The notion of *minimal rationality* is more realistic :

Minimal General Rationality Condition: If A has a particular belief-desire set, A would undertake some subset of the set of appropriate actions.[12]

Such a condition is dependent upon the *minimal inference* condition that if an agent had a particular belief-desire set, then that agent would make some, but not necessarily all, of the sound inferences from the belief set that are appropriate[13]. Such an ability is based on more realistic heuristic and deducing conditions on undertaking and performing *some* of the sound inferences from the belief set that would be appropriate for the agent to make[14]. Stringent conditions of rationality demand *ideal*

[10] Ibid., page 7.
[11] Ibid., page 13.
[12] Ibid., page 9.
[13] Ibid., page 10.
[14] Ibid., page 11.

consistency: if *any* inconsistency arose in the agent's belief set, the agent would eliminate it[15]. The minimal consistency condition requires that the agent would eliminate *some* of these inconsistencies *if* they were recognized. Inconsistencies in a belief set may exist simply because the logical relations among the beliefs may be not clear and hence not recognized, another source might be the *organization. storage and usage of those beliefs.*

The Structure of Human Memory

Cherniak argues that the understanding the structure of human memory is indispensable to an understanding of rationality or human capacities for rational behavior: actual models of human memory point to what our model of rationality should be. One such model rejects (what Cherniak considers) the extreme holism of Quine's belief web as being unrealistic. Not all of our beliefs are relevant to one another: the reevaluation of one belief does not necessarily lead to a reevaluation of all other beliefs. The web of belief model does not take into account the basic organization of human memory. For Cherniak, the web of belief is organized into a collection of "relatively independent subsystems". Connections are made less often (and are less likely) between these subsets[16]. A similar departure from the Quinean model can be noted in the following example:

> At least a decade before Fleming's discovery of penicillin, many microbiologists were aware that molds cause clear spots in bacteria cultures. and they knew that such a bare spot indicates no bacterial growth. Yet they did not consider the possibility that molds release an antibacterial agent..what makes this example philosophically significant..is the way we can explain it..we can say that the belief that molds cause bare spots seems to have been filed under the category of practical laboratory lore as information on undesirable contamination; the belief that a bare spot suggests inhibited growth seems to be in a different file; the microbiological

[15] Ibid., page 17.
[16] Ibid., page 51.

theory[17]

We can also dispose much of the associated paradoxicality of the Preface Paradox by pointing out its reliance on an idealization of the Quinean model: the size of the belief set for which the person makes a confession of error ("some of my beliefs are false") is crucial in assessing whether the person has made an accurate statement. The larger the belief set, the more likely the person is making an accurate statement and the more excusable the existence of the inconsistencies. This can be resolved by modifying the Quinean model. For our present purposes, the organized-into-subsets model of human memory best explains why a belief set is likely to be inconsistent: inconsistencies between elements in different subsets are less likely to be detected.

We can retain Quine's model by considering the web of beliefs to be a patchwork quilt with threads from adjoining patches running over into one another. All our beliefs are interconnected in that they are made from the same material. We can think of the interconnectedness of beliefs as arising from the fact that they are held by one agent and are dependent on some base conceptual scheme–their totality *is* the agent's conceptual scheme. Regardless, given our cognitive limitations, we treat beliefs as having a range of relevance (from utterly irrelevant to highly relevant) to each other. Beliefs about the composition of chemical compounds are highly relevant to beliefs about the efficacy of certain drugs in medical treatments and utterly irrelevant to beliefs about flexible exchange rates for currencies. *Bombay is hot* is irrelevant to *London is cold* but *Bombay is hot* is relevant to *I dislike the heat in West India.*

What model of human memory explains minimally rational performance? Firstly, only *a small subset of the total system should be activated at any given time.* Secondly,

[17]Ibid., page 50.

the contents of memory are *organized.* Items in memory are located for retrieval not by an exhaustive search but by a narrower search that exploits the structure of the memory system. Memory is represented as a sort of filing system where files contain subfiles[18]. A structure that would facilitate a narrower search would be one that allows searches to be confined to subsets of the total belief set. The relevance of beliefs within a subset would be a constraining factor in restricting searches to a particular subset: *searches would be regulated by relevance.* An example of such a compartmentalization is the example of Smith (an unfortunate agent for whom the compartmentalization fails as a good strategy):

> ..the belief that a flame can ignite gasoline is filed under roughly, "means of ignition"; the belief that the match he now holds has a flame is filed under "means of illumination."[19]

Why not just explain it as a change of opinion? Perhaps Smith does not really think that, at the time he lights the match, the match has a flame or that a flame can ignite gasoline. This entails our thinking of Smith as being a particularly indecisive agent. Smith is more plausibly described as an agent who carries out the compartmentalization described above rather than as one that is constantly changing his mind about what he believes.

Relevance, Belief Set Sizing and Degree of Compartmentalization

An adequate recall mechanism for beliefs requires a satisfactory *partial*, as opposed to an *exhaustive* search strategy; the search should be a *full search of only a subset of*

[18]Ibid., page 53. Models such as these are the semantic or neo-associationist models and constructive models. Most parallel-distributed approaches to human memory also presume the existence of a large number of memory/processing units working together rather than one simple monolithic entity [71]:

> Remembering...takes place in a parallel distributed processing system-a system consisting of a large number of simple but massively interconnected processing units

[19]Ibid., page 57.

all items in memory. The storage strategy itself should facilitate such a search (such a storage and search strategy should also explain minimal change in beliefs). Stored subsets or compartments should be organized according to subject matter, where items within a subset are more likely to be relevant to each other than items from different subsets[20]. Which of a given set of beliefs and desires should be grouped together will be dependent on what the agent's objectives are and on how much relevance the agent takes those beliefs to have to one another. Two items will be stored in different subsets or compartments if they tend not to be recalled together.

There are two ways in which one belief system might be more compartmentalized than another. First, one might have more compartments than another. Second, although they might have the same number of compartments at first, in one system, the compartments would be more permeable, leading in time to a merger of some compartments with others. Another parameter for compartmentalization is an *optimal sizing* of these subsets: too large and they quickly become useless for any kind of reasoning; too small and the chance of selecting the wrong subset and therefore missing a desired item becomes too high. To have belief subsets that are too small is to be in a cognitively undesirable position, that of being someone that must make new connections all the time.

One effect of the compartmentalization is that inconsistencies that involve beliefs in different subsets might go unrecognized[21]. When inconsistencies are unrecognized, beliefs from partitions might be used in successive inferences that lead to contradictory conclusions. We might show ourselves as holding on to inconsistent beliefs. To sum up:

[20] Ibid., page 66.
[21] Ibid., page 67.

> We have found the connection between memory organization and rationality: a basic condition for minimal rationality is efficient recall, which itself requires incomplete search, which in turn requires compartmentalization[22].

In conclusion, a plausible model of belief storage, recall and revision will be based on compartmentalization and will entail search and revision procedures that are computationally tractable. In doing so, it will ignore the ideal rationality, deduction, inference, consistency and heuristic requirements in favor of their minimalist versions.

What is common to the themes presented above (and what makes them plausible in that they do justice to *some* of our intuitions about an aversion to inconsistency) is that inconsistency, is hidden away or shielded and yet potentially accessible in certain acts of reasoning. The Cherniak model can be thought of as approximating the *corpus* model, in which not all the information in the corpus (a data base, an encyclopedia for example) is correct[23]. In the corpus model, different fragments[24] of the corpus come into action in different situations with the whole system of beliefs never manifesting itself all at once: the inconsistent beliefs are in separate *belief compartments*. The fragmentation of the corpus keeps it from appearing all at once. It often appears as one consistent corpus and then as another. The conflicts noticed between these fragments are the global inconsistencies of the corpus. Typically, trouble with such conflicts is avoided by not reasoning from mixtures of fragments (but, we might have

[22]Ibid., page 69.

[23]"Misinformation may render the corpus inconsistent", David Lewis, [68].

[24]Truth according to the corpus means the truth according to some one fragment. There is no guarantee then, that implication preserves truth according to the corpus, unless all the premises are drawn from a single fragment. Conclusions drawn from two or more premises taken from disagreeing fragments may be true according to no one fragment and therefore are not true according to the corpus. Such a definition of truth within such a corpus gives us a motivation for a four-valued semantics for logics such as those proposed by Dunn [19], Belnap [5], Ginsberg [40] and Fitting [24, 26]. In these logics, there are ***truth value gaps*** (statements that are neither true or false) and ***over definition*** (statements that are true and false). We will employ precisely such a logic in our B-structures model.

to at times).

Cherniak's model serves as an amplification of points made earlier where we provided arguments against the logical closure and consistency conditions on belief representations. In adopting models based on this approach, we aim to stick to the middle ground between a purely prescriptive approach that fails to recognize the natural limitations of human agents and a purely descriptive approach that pays no attention to any plausible intuitions about what form rational reasoning and belief change should take. Our model for reasoning is an inconsistency tolerant one: one in which inconsistencies in a corpus are introduced by fragmentation of beliefs. These fragments can *overlap* and the inconsistent conjunction of fragments is in and of itself not true according to any one fragment. The agent is not implicated in any inconsistencies since it *typically* reasons from or "activates" one fragment at a time.

Chapter 3

Paraconsistency: Logics and Belief Revision

I now briefly discuss some approaches to modeling inconsistency via paraconsistent logics and the application of these logics to belief revision. The first such treatment occurs in the systems of *relevance logics* of Anderson and Belnap [3]. The logics concern themselves with the 'paradoxes' of strict and material implication. Examples of problematic material implication are:

- $p \rightarrow (q \rightarrow p)$.
- $(p \rightarrow q) \vee (q \rightarrow r)$.

Examples of problematic strict implication are:

- $(p \wedge \neg p) \rightarrow q$.
- $p \rightarrow (q \rightarrow q)$.

In each case above, there antecedent appears irrelevant to the consequent. Anderson and Belnap present a system of natural deduction that blocks such problematic implications by imposing (amongst other conditions) the condition that the antecedent and consequent must share propositional variables. This system also blocks inferences such as "Grass is made of steel. Therefore, either I am male or I am not" which classical logic would permit. The sharing of variables ensures that only relevant premises play a role in the derivation of conclusions.

The two models that I consider next are in the non-adjunctive tradition of paraconsistent systems and have something in common: they take the treatment of inconsistencies to be solvable by a suitable partitioning of the total belief set. Such a partitioning coupled with a rejection of the semantic principle of adjunction (in the case of Rescher and Brandom) and the syntactic rule of adjunction (in the case of Schotch and Jennings) blocks the inference of arbitrary propositions from a possibly inconsistent set. These logics correspond to the corpus model referred to in the previous chapter (only partially. since in these models there are no overlapping fragments of the corpus). An important question remains unanswered: *what makes the belief set structured in such a manner?* Our suggestion is a division of beliefs on the basis of relevance. Such an answer is implicit in the approaches; our remaining task. after studying the details of these approaches, is to give a formal definition of relevance. This will aid us in the definition of an inference operation from inconsistent beliefs that has relevance sensitivity built into it.

3.1 Rescher and Brandom's Logic of Inconsistency

The logic of inconsistency provided by Rescher and Brandom (**RB**) [89] provides a semantics for non-standard worlds that rejects the semantic principle of adjunction towards a resolution of the logical paradoxes[1].

[1]This approach is also referred to as the *impossible worlds* approach and is adopted by [12. 87. 56]. In general, in impossible worlds, $p \wedge \neg p$ might hold.

Standard possible worlds for **RB** are those in which truth valuations are available for every proposition (i.e., the law of the excluded middle holds) and for any proposition, either it or its contrary obtains in that world (i.e., bivalence holds). In non-standard possible worlds, the law of the excluded middle fails for *schematic* worlds and bivalence fails for *superposed* worlds (definitions below). A world description is an assertion that claims certain things to be the case. Proposition α is true in world w if the ontological state of affairs corresponding to its truth obtains, that is, $t_w(\alpha)$ iff $[\alpha]_w = True$. **RB**'s perspective is ontological, rather than reflecting an epistemic concern with our knowledge, information or beliefs about this world[2]. The non-standardness at issue characterizes the world itself. Inconsistent worlds are constructed from standard worlds by the processes defined below (the resulting logic has truth-value gaps):

Definition 5 [89] *$[\alpha]$ is true in the schematized world $w_1 \cup w_2$ (i.e., $t_{w_1 \cup w_2}(\alpha)$) iff $[\alpha]_{w_1} = True$* **and** *$[\alpha]_{w_2} = True$ (i.e., $t_{w_1}(\alpha) \wedge t_{w_2}(\alpha)$)*

Definition 6 [89] *$[\alpha]$ is true in the superposed world $w_1 \uplus w_2$ (i.e., $t_{w_1 \uplus w_2}(\alpha)$) iff $[\alpha]_{w_1} = True$* **or** *$[\alpha]_{w_2} = True$ (i.e., $t_{w_1}(\alpha) \vee t_{w_2}(\alpha)$)*

Now, some standard semantical principles (such as $t_w(\alpha)$, $\alpha \vdash \beta \Rightarrow t_w(\beta)$ for all w; $\vdash$ is classical entailment) still hold, but others such as $t_w(\alpha), t_w(\beta) \Rightarrow t_w(\alpha \wedge \beta)$ fail. **RB** retain the syntactic rule of adjunction $\alpha, \beta \vdash \alpha \wedge \beta$ but reject the principle of semantical adjunction $t(\alpha), t(\beta) \Rightarrow t(\alpha \wedge \beta)$ since it fails for non-standard worlds. For example, let $t_{w_1}(p \wedge \neg q)$ and $t_{w_2}(\neg p \wedge q)$ both hold; in the superposed world $w = w_1 \uplus w_2$ both $t_w(p)$ and $t_w(q)$ obtain and yet, we do not have $t_w(p \wedge q)$. We have

[2] [89], page 5.

an unorthodox semantics while orthodox rules of inference continue to hold. **RB** propose a distributive and a collective reading of the principle "valid inferences from true premises yield true conclusions". The distributive rule is:

Distributive Rule: (DR) Whenever $\alpha_1, \alpha_2, ...\alpha_n \vdash \beta$ is a valid inference principle of classical logic, and $t_w(\alpha_1), t_w(\alpha_2), t_w(\alpha_3), \ldots t_w(\alpha_n)$ then $t_w(\beta)$.

The collective reading is:

Collective Rule: (CR) Whenever $\alpha_1, \alpha_2....\alpha_n \vdash \beta$ is a valid inference principle of classical logic, and $t_w(\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \ldots \alpha_n)$ then $t_w(\beta)$.

RB suggest that DR fails for all superposed worlds. One can retain classical logic in non-standard worlds, by only using CR instead of DR. The *distinction to be made here is between the conjunction $\alpha \wedge \beta$ and the juxtaposition $\alpha.\beta$*, a distinction that is commonly made in probabilistic and inductive reasoning (see Kyburg, [60, 62]). **RB** argue that any semantic principle which involves the combining of distinct premises (any rule that uses a series of premises distributively) in a superposed world can possibly fail[3]. The failure of semantic principles involving collective-truth readings is crucial in the **RB** system's treatment of Russell type paradoxes. These paradoxes have a general form: given a set of basic assumptions, $\alpha_1....\alpha_n$, we derive both X and $\neg X$. The problematic assumption is some α_i. Normally, one would drop or replace α_i by some other assumption α_i^*; **RB** suggest a *division*: $\alpha_i = (\alpha_i^1 \wedge \alpha_i^2)$. This division means, now, that the replacement of α_i within the assumption set by either of these two sets produces a consistent result:

> Such a resolution would enable us to view the initial assumption-set $\alpha = \{\alpha_1, \alpha_2, \alpha_3...\alpha_n\}$ as the *collectively inconsistent* superposition of the two discordant assertion-zones at issue. Our set-theory becomes an inconsistent whole with local irregularities located in the environment of prob-

[3]This leads, for example, to a failure of the semantic principle of *modus ponens* as well: If $t_w(\alpha)$ and $t_w(\alpha \supset \beta)$ then $t_w(\beta)$.

> lematic axioms whose problematic impact could be removed by a suitable *splitting* . On this approach, the proper treatment of inconsistency proceeds not by the revision of an inconsistency-producing axiom, but by its division.[4]

The problems with **RB**'s suggestion above is that it is unclear *how an assumption is to be divided.* There is a slight danger here that their suggestions stay at the level of metaphor[5].

To use this semantics for an inconsistent theory[6] (as our theory of the world is) we must divide the axioms of the theory. By dividing the axioms into two sets it may be possible to achieve *local consistency* by means of two sets of axioms that by themselves are consistent. So, for example, our inconsistent theory might consist of $\{p, q, \neg p\}$ but, a division would result in either $\{\{p, q\}, \{\neg p\}\}$ or $\{\{p\}, \{q, \neg p\}\}$ i.e., two consistent axiom sets. Both p and $\neg p$ belong to the composite theory but no arbitrary propositions belong to the theory (thus taking care of the problem of explosiveness). For our purposes, of importance is the notion of adding up an agent's beliefs to arrive at an inconsistency. Locally, no such inconsistency is to be found. Extending their apparatus to a modeling of an agent's beliefs, **RB** conclude:

> With our apparatus, we can express...contextual relativity of belief, representing the various contexts by *different subsets of beliefs of the speaker* about the circumstances of the discourse, and the different dispositions to assent as the results of *conjunctive multi-premise inferences* from those beliefs [7]

[4]Ibid., page 37.

[5]See Dale, [15] for a critical assessment of the **RB** approach to handling the set-theoretic paradoxes.

[6]In this paragraph, I depart from the conventional usage of the term 'theory' as being a closed set of sentences. Here 'theory' simply means a set of beliefs.

[7][89], page 101.

3.2 The Schotch and Jennings Model

Schotch and Jennings' [98] (**SJ**) model drops the adjunction rule in order to obtain an entailment relation suitable for reasoning with inconsistent sets. Taking the position that consistency is an "ideal state of affairs often approximated rather than achieved", **SJ** ask the following question:

> what principles of inference are suitable for reasoning from inconsistent premises?[8]

In response, a general notion of inconsistency, one that admits of *levels* or *degrees* is introduced–the classical notion of consistency emerges as a special case[9]. **SJ** employ the idea of a *level of coherence.* Consider a finite set, Γ, and consider a partition which is a family of disjoint subsets, each of which is classically consistent and whose union is Γ. A coherence function c, for Γ, is a function with domain the set of all finite sets of sentences, co-domain the set $Nat \cup \{\omega\}$ (Nat is the set of natural numbers), and is defined as follows:

Definition 7 [98] *For $\bot \notin \Gamma$, $c(\Gamma) = m$ where m is the least integer such that Γ can be partitioned into m sets $a_1, a_2, a_3 \ldots a_m$ all of which are individually consistent: $a_i \nvdash \bot$ $(\forall i \leq m)$.*

If $\bot \in \Gamma$ then **SJ** adopt the convention that $c(\Gamma) = \omega$ and define a relation between finite sets of sentences and sentences written $\Gamma [\!\vdash \alpha$:

Definition 8 [98] *for $c(\Gamma) = n(\omega)$, $\Gamma [\!\vdash \alpha$ iff for every n-fold (ω-fold) decomposition of Γ, $a_1 \ldots a_n$, there is some i such that $a_i \vdash \alpha (1 \leq i \leq n(\omega))$*

[8] [98], page 328.
[9] In the case of B-structures, a similar notion emerges.

As with the standard notions of entailment $\vdash$, $[\vdash$ is determined by rules of inference[10] which are introduced by first defining a m- cluster:

DEFINITION 9 [98] *For any finite set a, $C \subseteq 2^a$ is an m-cluster iff: $m \in Nat$ and* $(\forall f \in m^a, \exists x \in C, \exists y \leq m : x \subseteq f^{-1}[y])$

That is, for any way of dividing a into m subsets there is a member x of C such that x is included in at least one of the m subsets into which a has been divided. Or, a set of sets of sentences $\{\{\alpha_i, \alpha_j, \ldots, \alpha_m\}, \ldots, \{\alpha_n, \ldots, \alpha_p\}\}$ drawn from $\alpha_i, \ldots, \alpha_q$ is an m-cluster iff the set $\{\{i, j, \ldots, m\}, \ldots, \{n, \ldots, p\}\} \subseteq 2^q$ is an m-cluster. A structural rule is constructed for $[\vdash$ which depends upon the coherence level of a set Γ:

DEFINITION 10 (C) *If $C = c_1, c_2, \ldots, c_n$ is an m-cluster constructed out of $\alpha_1, \ldots, \alpha_k$ $\in \Gamma$ and $c(\Gamma)=m$, then*

$$\frac{c_1 \vdash \beta, c_1 \vdash \beta, \ldots, c_1 \vdash \beta}{\Gamma [\vdash \beta}$$

From this definition, it follows that one classical operator rule is lost: from Γ $[\vdash \alpha$ and Γ $[\vdash \beta$ it no longer follows that Γ $[\vdash \alpha \wedge \beta$. This is now true only in the special case that $c(\Gamma) = 1$. Each of the classical rules which fail in general all hold for the special case of $c(\Gamma) = 1$ i.e., when Γ is consistent.

Theories in Schotch and Jennings

The structural rule defined for $[\vdash$ above also enables a more general definition of *theories*. A restrictive concept of theories is defined via classical rules of inference:

DEFINITION 11 [98] *Δ is a theory iff for every sentence α, $\Delta \vdash \alpha \Rightarrow \alpha \in \Delta$.*

[10] These consist of structural rules (similar to those provided for the classical notion of $\vdash$ such as $\alpha \in \Gamma \Rightarrow \Gamma \vdash \alpha$) and operator rules. For example, an operator rule for the classical $\vdash$ is:

$$\frac{\Gamma \vdash \alpha, \Gamma \vdash \beta}{\Gamma \vdash \alpha \wedge \beta}$$

For example:

$\alpha \in \Delta \ \& \vdash \alpha \rightarrow \beta \Rightarrow \beta \in \Delta$

$\alpha \in \Delta \ \& \ \beta \in \Delta \Rightarrow \alpha \wedge \beta \in \Delta$

We can now define m-theories:

DEFINITION 12 [98] *Δ is a m-theory iff $c(\Delta) = m$ and $\Delta[\vdash \alpha \Rightarrow \alpha \in \Delta$*

The older, classical concept of a theory is replaced by the definition above and the closure condition of 1-theories turns out to be a special case of a more general condition:

DEFINITION 13 [98] *Δ is an m-theory iff:*

($\mathcal{A}$). $\alpha \in \Delta$ & $\vdash \alpha \rightarrow \beta \Rightarrow \beta \in \Delta$

and:

($\mathcal{B}$) $c_1, \ldots, c_k \subset 2^{\Delta}$ is an m-cluster $\Rightarrow \{\wedge c_1, \ldots \wedge c_k\} \in \Delta$ when "$\wedge c_1$" denotes "$\alpha_i \wedge \ldots \wedge \alpha_{ij}$" for $c_i = \alpha_{i1}, \ldots, \alpha_{ij}$

When $m = 1$, any finite subset of Δ is a an m-cluster so that $\mathcal{B}$ reduces to the standard closure condition. For **SJ**, belief sets and sets of sentences which ought to be true, if they form theories, do not form 1-theories; they form m-theories with the rule of inference $[\vdash$ defined above. **SJ** achieve a partitioning of the belief set by the notion of maxi-consistency, and take inference from such a cluster to be regulated by their new structural rules. Since each set is maxi-consistent, the entire cluster could be inconsistent and yet, not be trivial.

Example: Consider the set $\{p, \neg p, q, r\}$. This has two maximally consistent sets. $\{p, q, r\}$ and $\{\neg p, q, r\}$. In addition, this set has level 2 since the least number of partitions we can divide it into which are individually consistent is 2. We can split it into

$c_1 = \{p,q,r\},\{\neg p\}, c_2 = \{p,q\},\{\neg p,r\}, c_3 = \{p,r\},\{\neg p,q\}$, or $c_4 = \{\neg p,q,r\},\{p\}$. We cannot derive a contradiction from this set, since no contradiction is implied by either of its partitions. Now $q \wedge r$ does not follow since there are decompositions c_2, c_3 above, in which neither of the two sets imply it. Priest [82] has termed this a counterintuitive result. However, if we are making a distinction between the presence of two conjuncts and the conjunction then this is not all that implausible.

3.3 Paraconsistent Belief Revision

The possibility of paraconsistent belief revision has been explored by Graham Priest [83]; in the previous chapter, we considered some of Priest's arguments for a model of belief change that renders it intelligible that an agent whose beliefs were inconsistent would still be interested in revision. Koji Tanaka [101] and Restall and Slaney [90] have developed paraconsistent semantical representations based on the revision of models approach suggested in [44]. Restall and Slaney use *first-degree entailment* (see Dunn, [19]) as the consequence relation for the underlying logic (formulae take as truth values subsets of $\{T, F\}$ rather than just T, F alone) associated with the belief revision operations. In first-degree entailment a formula α is entailed by a set of formulas X if and only if any valuation that assigns to every element of X either $\{T\}$ or $\{T, F\}$ also assigns to α either $\{T\}$ or $\{T, F\}$.

There are some positive results of this application of first-degree entailment to belief revision: many of the powerful representation theorems associated with constructions based on entrenchment and systems of spheres modelling are obtained. There are some interesting failures as well: the axiom of recovery fails as does the Harper Identity. In the case of the Harper Identity it becomes possible to have $\alpha \in K^{\dot{-}}_{\alpha}$

since α might be both in the original set and in the set K revised by $\neg\alpha$.

Tanaka's paraconsistent model of belief revision offers three different semantics for belief revision based on systems of spheres:

- Relevant logic systems (Anderson and Belnap, [3], see discussion above).

- Positive-plus systems (da Costa [13]). The central intuition in these systems is to maintain the apparatus of classical logic, but to devise a non truth-functional notion of negation in an interpretation. Take an interpretation to be a function which maps formulas to 1 or 0. Now, $\wedge, \vee, \rightarrow$ behave classically, but the value of $\neg\alpha$ is taken to be independent of that of α. Both, for instance, may take the value 1. Negation has no significant properties under such a semantics. Various properties of negation may be obtained by adding further constraints on interpretations. If for any α, either α or $\neg\alpha$ must take the value 1 (Law of Excluded Middle) and whenever $\neg\neg\alpha$ takes the value 1, so does α, we get da Costa's systems C_i for finite i.

- Non-adjunctive systems of discoursive or discussive logic (Jaskowski, [57]). In a discourse, each participant contributes some information or opinions. Truth in a discourse is the sum of the participants contributions. Each participant's opinions are (assumed to be) self-consistent, but are possibly inconsistent with those of others. Formally, take an interpretation I for a standard modal logic such as $S5$. Each participant's belief set is the set of sentences true in a possible world in I. Thus, α holds in I iff α holds at some world in I. Clearly, one may have both α and $\neg\alpha$ (but not $\alpha \wedge \neg\alpha$) holding in an interpretation. Since *modus ponens* for the material conditional fails, Jaskowski introduces a connective called discussive implication, $\supset_d$, defined as $\diamond(\alpha \supset \beta)$. It can be shown that in $S5$ discussive implication satisfies *modus ponens*.

The systems of spheres that are obtained by Tanaka have some interesting properties. Most notably:

- All three systems violate the recovery postulate.
- Axiom (K^*5), the axiom of Vacuity is violated.
- Axiom (K^*7), the axiom of Superexpansion is violated.
- The soundness proof for systems of spheres based on relevance logics does not need the axiom $K^{\dot{-}}4$ i.e., beliefs that are logical truths occur in all belief sets. Yet, as Tanaka correctly points out, this is not needed since we can well imagine rational human beings that do not believe all logical truths, *qua* logical truths. Intuitionist logicians might not believe in the logical truth of $\alpha \vee \neg\alpha$ for instance. In belief revision based on relevant logics, it becomes possible for any belief to be dropped from the belief set, even logical truths.

In general, the loss (in terms of plausibility or realistic revision operators) in going to paraconsistent models for belief revision is not significant. We essentially lose the restrictions that arise because of the proscription of inconsistency: the counterintuitive axioms of Recovery and Superexpansion. We have already provided arguments against Superexpansion in Chapter 1; for arguments against the Recovery postulate see Hansson [46] and Levi [66]. The loss of the axiom ($K^{\dot{-}}4$) (Preservation) is more arguable (though Ryan [95] has provided plausible counterexamples against it) but certainly, none of the dire consequences predicted by allowing inconsistency seem to hold.

Conclusion

There are two morals that I draw from the preceding discussions: the need to partition beliefs in such a way that inconsistencies can be quarantined, and secondly, the need to play close attention to problematic rules of inference such as the rule of adjunction that lead to the vacuity of a set containing an inconsistency. Furthermore, the only loss in going to paraconsistent models of belief revision is the loss of rationality postulates against which an independent case for rejection can be made. One way then, to tolerate inconsistencies in the belief representation could be a form of partitioning and a reasonable candidate for the logic of inference from such a representation is one that loses some of the restrictions of classical logic (possibly a non-adjunctive system).

One issue remains untouched however: what is the most promising strategy to explain the partitioning of beliefs? Our contention is that it will be in terms of some relevance relation holding between beliefs that exist in the same subset. Not co-incidentally, one factor common to the trivial update, the proscription against inconsistency and problems with computational tractability is the persistent failure to deal with the notion of relevance relations amongst beliefs. This failure to explore the notion of relevance fully is also present in the approaches that we have considered thus far. Both the **RB** and **SJ** systems of non-adjunctive logic presume that a division of the premise set is available; both systems assume that the drive to attain consistent subsets is responsible for the division of the subsets. If a plausible description could be given of the role that relevance relations could play in the division of the belief representation, a motivated model for inconsistency tolerance could be provided. I now describe a model for belief revision that incorporates a notion of relevance sensitivity and points us in the way of how to handle the problems raised above. This will serve as the foundation for a model that will take the middle ground between paraconsistent and traditional logic constrained approaches. between belief set and belief base approaches.

Chapter 4

The Language Splitting Model

In this section I examine the Language Splitting (**LS**) model of beliefs (Parikh. [80]) which addresses the minimal change and tractability requirements raised in the preceding critique of the AGM model for belief revision.

4.1 The Language Splitting Model

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

DEFINITION 14 [80] *L_1, L_2 split the language L relative to a theory T iff there are formulae α, β such that $\alpha \in L_1, \beta \in L_2$, and $T = Cn(\alpha, \beta)$. Similarly, (mutually disjoint) languages $L_1, \ldots, L_n$ split T iff there exist $\alpha_i \in L_i$ such that $T = Cn(\alpha_1, \ldots, \alpha_n)$; $\{L_1, \ldots, L_n\}$ is a T-splitting.*

We can say that T is 'generated' by the T_i in the sub-languages L_i. If $L_1 \subset L$, then T is said to be *confined* to L_1 if $T = Cn(T \cap L_1)$. T, then, knows nothing about the part $L - L_1$ of L. The first part of the definition implies that T contains no 'cross-talk' between L_i and L_j for distinct i, j. The following lemmas are instrumental in proving the existence of a new update operator for theories split by language:

LEMMA 1 [80] *Given a theory T in the language L, there is a unique finest T-splitting of L, i.e., one which refines*[1] *every other T-splitting.*

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

LEMMA 2 [80] *Given a formula α and a language L, there is a smallest language L' (a set of propositional atoms) in which α can be expressed, that is, there is a $L' \subseteq L$ and a formula $\beta \in L'$ with $\alpha \Leftrightarrow \beta$, and for all L'' (another set of propositional atoms) and β'' such that $\beta'' \in L''$ and $\alpha \Leftrightarrow \beta''$, $L' \subseteq L''$.*

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

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

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

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

DEFINITION 15 *Formulas α, β are relevant to each other $\mathcal{R}(\alpha, \beta)$ iff their smallest languages L_α, L_β share propositional atoms i.e., iff $L_\alpha \cap L_\beta \neq \emptyset$*

Using the smallest language blocks trivial relevance between two formulas. For example, if we were not using the smallest language, the formulas $\alpha = p \wedge (q \vee \neg q)$, $\beta = r \wedge (q \vee \neg q)$ would be deemed relevant, but under our definition they are deemed irrelevant. We now clarify the notion of a language *compatible* with a given T-splitting:

DEFINITION 16 [80] *Given a theory T, a language L, and a formula α, let L_α be the smallest language in which α can be expressed and L^T_α be the smallest language containing L_α such that $\{L^T_\alpha, L - L^T_\alpha\}$ is a T-splitting.*

L^T_α is the smallest union of elements of the finest T- splitting of L in which α can be expressed.

The Language Splitting Axioms

Parikh's rationale for the language splitting axioms is as follows:

> If we have information about two subject matters which as far as we know are unrelated, [are split] then when we receive information about one of the two, we should only update our information in that subject and leave the rest of our beliefs unchanged.[2]

Such a condition obviously agrees with the Gärdenfors *preservation criterion*. This central intuition is expressed by the following (which supplements AGM axioms: in the statements of the axioms below, all operators are intended to be general AGM-compatible update operators):

[2] [80], page 3.

Axiom P: If $T = Cn(\alpha, \beta)$ where α, β are in L_1. L_2 respectively and γ is in L_1. then T updated by γ is $Cn(\alpha) \ominus \gamma \dotplus \beta$, where $\ominus$ is an update operator for the language L_1.

Three additional axioms for revision on theories split by language follow immediately:

Axiom P1: If T is split between L_1 and L_2, and α is an L_1 formula. then $T * \alpha$ is also split between L_1 and L_2.

Justification: The new information received makes no connection between the partitions in the theory, so after revision the old partition should be retained.

Axiom P2: If T is split between L_1 and L_2, α, β are in L_1 and L_2 respectively then $T * \alpha * \beta = T * \beta * \alpha$.

Justification: The order of the receipt of two unrelated pieces of information should make no difference to the revision of our beliefs.

A related axiom, supplementary to the original axioms (and compatible with them) has been suggested by K. Georgatos:

Axiom P2g: If T is split between L_1 and L_2, α, β, are in L_1 and L_2 respectively then $T * \alpha * \beta = T * \beta * \alpha = T * (\alpha \wedge \beta)$.

Axiom P3: If T is confined to L_1 and $\alpha \in L_1$ then $T * \alpha$ is just the set of consequences in L of $T \ominus \alpha$ where $\ominus$ is the update of T by α in the sublanguage L_1.

Justification: The agent possesses no information about $L - L_1$ and has not received any yet. If so, the agent should update as if it held beliefs only in L_1. $L - L_1$, about which we had no prior information and have not received any updates should not have any impact.

OBSERVATION 1 [80] *The trivial update*[3] *(as in Definition 3), cannot satisfy Axiom* **P2** *or Axiom* **P** *though it does satisfy* **P1** *and* **P3**. *Therefore, any update operator that satisfies Axiom* **P** *cannot be the trivial update.*

Demonstration: Let $T = Cn(p, q)$ and consider new inputs $\alpha = p$ and $\beta = \neg q$. Then the trivial update yields $T * \alpha * \beta = T * \beta = Cn(\neg q)$ and $T * \beta * \alpha = Cn(\beta) * \alpha = Cn(\neg q, p)$. This violates **P2**. Also $T * \beta = Cn(\neg q)$ which violates Axiom **P**. Therefore **P** or **P2** alone rules out the trivial update.

The following theorem provides a new update operator for theories on which splittings can be carried out:

THEOREM 4.1 [80] *There exists an update operator that satisfies Axiom* **P** *and axioms* $(K^*1) - (K^*6)$; *the update operator blocks the trivial update.*

One possible revision procedure is defined as follows[4]:

DEFINITION 17 [80] *Given* T, α *if* $\neg\alpha \notin T, T *_s \alpha = T \dot{+} \alpha$. *Otherwise* $T = Cn(\beta, \gamma)$ *where* $\beta \in L^T_\alpha, \gamma \in L - L^T_\alpha$ *respectively. Then* $T *_s \alpha = Cn(\alpha, \gamma)$.

Example: Let $T = Cn(p, q \vee r, s)$; the partition $\{\{p\}, \{q, r\}, \{s\}\}$, is the finest T-splitting. Let $\alpha = \neg p \wedge \neg q$. Then $L^T_\alpha = \{p, q, r\}, \beta = p \wedge (q \vee r)$ and $\gamma = s$. β represents the part of T incompatible with the new information α. Therefore $T *_s \alpha$ is $Cn((\neg p \wedge \neg q), s)$. The update procedure of the theorem notices that α is not in any conflict with s and retains it.

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

[3] If α is consistent with T, then $T * \alpha = T \dot{+} \alpha$, otherwise $T * \alpha = Cn(\alpha)$.
[4] It is obvious that other revision procedures exist.

> ..when we get a piece of information which lies in one of the sub-languages (or straddles only two or three of them) then we can leave most of the theory unchanged and revise only the affected part.[5]

We now have a notion of tractable, localized, relevance-sensitive belief revision to work with. The minimal change requirement is satisfied by considering as small a part of the old theory as possible as a candidate for change. Considering only a part of the original belief set ensures that while the complexity of the revision procedure stays at the same level, the size of the problem instance is drastically reduced.

Remarks:

- The notion of splitting languages and the lemmas can be extended to first order logic without identity[6]. Arguments similar to those in [80] can be provided to show the existence of an update operator satisfying AGM axioms and Axiom **P** for theories in first-order languages.
- The **LS** model is not bound to a particular assignment of propositional atoms to the language. As an example, the language $L = \{p, q\}$ is also generated by r, q where r is $p \leftrightarrow q$. This follows immediately from the fact that p can be expressed in terms of q, r as $p \Leftrightarrow (q \leftrightarrow r)$.

4.2 Relevance and Knowledge Representation

The **LS** model has provided us a with a relevance sensitive model for belief revision. It is this sensitivity that we will aim to exploit in the models that we develop. Is the relevance relation provided adequate to capture the notion of cognitive relevance? The problem of defining a notion of relevance between propositions is notoriously tough.

[5][80], page 5.

[6]The introduction of identity introduces problems pointed out by Sam Buss.

One reminder of how difficult this problem is is to remember that defining a notion of relevance that is adequately formalizable is part of a possible resolution of the infamous *frame problem* in artificial intelligence. In this problem given (an adequate representation of) a state of a system, and the post-condition of an action performed in the state, what is the state resultant upon the performance of that action? The post-condition should be true in the resultant state and change as little of the old state as possible. An adequate solution to the frame problem (for belief states) consists in explaining how it is that our beliefs rest on certain default assumptions. how this frame of beliefs has a certain resiliency and how is it that changes in our beliefs do not seem to necessitate widespread changes in the frame. Our problem of minimal change of beliefs is thus neatly framed (no pun intended) in terms of the frame problem. The solution that the language splitting model suggests is that the only changes made during belief revision are to formulas that are in the same language or have language overlap with the new information. This corresponds to the *sleeping dog strategy* suggested by McDermott [70] as a solution for the frame problem:

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

The sleeping dog strategy depends on a particular representation of the elements in the knowledge base under consideration:

> [If] you want to use a sleeping dog algorithm for your database, you must first devise a system of canonical representation for the facts.[7]

Part of the problem of providing a canonical representation for the facts is the problem of packing into the representation the relations between the facts such that the

[7] Fodor, [27], page 140.

relevance relation between the facts can be 'read off' when the time comes to make a change to the database. If we had a precise definition of relevance (one that formalized the distinction between relevant facts and irrelevant ones), then we would have a basis for an implementation of the sleeping dog strategy. It is this precise method of course, that is notoriously problematic.

A straightforward rejection of the idea that relevance can be defined in a propositional language is made by Gärdenfors:

> [A] criterion of this kind cannot be given a technical formulation in a model based on belief sets built up from sentences in a simple propositional language because the notion of relevance is not available in such a language[8].

A notion of relevance that captures *all* the different nuances of the term as used in reasoning probably cannot be provided in a propositional language. It is unclear how it could be provided in a first-order one either: in order to express relevance precisely, one would want a language in which we could quantify over relations. Perhaps one should make do with relevance relations that are represented explicitly and are easy to compute so as to lend themselves to a feasible implementation. A relevance relation that cannot be 'read off' easily from a database is not likely to lead to a tractable solution. If we did have a relevance relation that could be 'read off' easily then that definition of relevance would create a situation in which the organization of beliefs would create particular belief revision strategies. So our beliefs would be organized by relevance and therefore searches and belief revision which are a function of this organization, would be regulated by relevance as well. The reason the sleeping dog strategy would work in belief change is that that would be just how beliefs were best organized and revised.

[8] [33], page 351.

4.2.1 Epstein's Relatedness Logic

Richard Epstein [22] offers a *logic of relatedness*[9] between propositions which offers some interesting intuitions that are captured by the notion of language overlap. There are some assumptions at the heart of this treatment. Firstly, the subject matter of a proposition is independent of its truth value. Secondly, a proposition can be viewed as being related in subject matter to some propositions and unrelated to all others. Thirdly, the subject matter of a proposition is more dependent on context and more holistic than truth. As an example, 'London is foggy' is relevant to 'Delhi is hot' in some contexts, say *climate in world capitals*, and yet in some contexts it is irrelevant to 'Delhi is hot' as in deciding whether to go for a afternoon walk in the summer in Delhi. Epstein places the following conditions on an arbitrary relevance relation $\mathcal{R}$ (for propositions α, β):

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

Epstein intends the above as conditions on *any* plausible notions of relevance; some are more intuitive than others (as an example, it is not clear what intuition about relevance **R2** aims to capture). Notice that the above list does not include transitivity: including transitivity would quickly trivialize relevance. However, a weaker version of

[9]Epstein's treatment appears to try and formalize the relation of relevance. In this section therefore, we will stick to the term 'relevance'.

transitivity can be developed (and will be presented in the section on belief sequences). In the axiomatization above, relevance is assumed to be a binary relation. In the section on sequences, a ternary relevance relation will be developed. **R1-R5** are incompatible with the following:

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

We make a distinction here between $L(\alpha)$, the set of propositional variables in a formula α and L_α as in Definition 15 and use this to distinguish between two possible relevance relations based on language overlap:

DEFINITION 18 *α, β are related by syntactic language overlap ($\mathcal{R}_s(\alpha, \beta)$) if $L(\alpha) \cap L(\beta) \neq \emptyset$. α, β are related by logical language overlap ($\mathcal{R}_l(\alpha, \beta)$) if $L_\alpha \cap L_\beta \neq \emptyset$.*

It is obvious that $\mathcal{R}_l(\alpha, \beta))$ implies $\mathcal{R}_s(\alpha, \beta))$. Rodrigues [91] has shown that the relation ($\mathcal{R}_s(\alpha, \beta)$) is the smallest relation satisfying Epstein's conditions. However:

OBSERVATION 2 *The relevance relation $\mathcal{R}_l(\alpha, \beta)$ using the smallest languages for α, β satisfies conditions* **R1, R3-4, R6** *above. It does not satisfy conditions* **R2, R5**.

Subject Matters

The definition of relevance provided in the **LS** model let us define the subject matter of a proposition as the language of the formula used to express the proposition. This lets us speak of subject matters being identical or overlapping[10]. The languages of propositions can overlap, the language of a proposition can be contained in the language of another and can be identical with that of another. That is, for propositions α, β, if $L_\alpha = L_\beta$ then α, β are about the same subject; if $L_\alpha \cap L_\beta \neq \emptyset$

[10] Relevance between propositions can be cast in probabilistic terms as well i.e., α, β are relevant iff $P(\alpha|\beta) \neq P(\alpha)$. In this case, two propositions could be relevant even if there is no language overlap in the formulas.

then the two propositions have some subject matter in common and if $L_\alpha \subseteq L_\beta$ then the subject matter of α is contained in the subject matter of β.

Wasserman on Relevance

I now briefly describe an alternative, primarily logical, definition of relevance between propositions. Wasserman [106] defines the relevance relation $\mathcal{R}$ on a belief base H for propositions ϕ, ψ as follows:

DEFINITION 19 [105] *$\mathcal{R}(\phi, \psi)$ if and only if there is a set $X \subseteq H$ such that $X \cup \{\phi\} \nvdash \bot$ and either ($X \nvdash \psi$ and $X \cup \{\phi\} \vdash \psi$) or ($X \nvdash \neg\psi$ and $X \cup \{\phi\} \vdash \neg\psi$).*

This relation is not symmetric. However we would expect that if a formula was relevant to another, it is because of the relation that the *two formulas bore to one another*. Under what circumstances could we take α to be relevant to β and yet consider β to be irrelevant to α? Perhaps α is relevant to considerations of β by virtue of scale, whereas β is not relevant to considerations of α by that same scale? Another could be that often symmetry of relevance only arises in particular contexts. An example (due to Parikh): a change in tax law may affect my standard of living but a change in my standard of living is unlikely to bring about any change in the tax laws.

Wasserman notes that calculating related pairs above could be very difficult. If, however, the relation was given, then calculations could be made much simpler as follows: list all sets of formulas. For each pair of formulas ψ, ϕ select all sets in which ψ occurs, check whether they imply ϕ or $\neg\phi$. If they do, then check each subset to see whether they imply ψ or $\neg\psi$. If they do not imply ψ or $\neg\psi$, we have $\mathcal{R}(\psi, \phi)$. This is an extremely expensive procedure since derivability needs to be checked several times over. If however, the same language is to be used, the relation can be pre-computed and used for all belief sets.

In Defense of Language Overlap as a Relevance Relation

The notion of relevance based on smallest language overlap that we have worked with in this treatment is a hybrid notion: based on the syntactic form and logical content of the formulae in question. Is such a notion of relevance is too simplistic? Perhaps, but to offer an adequate account of relevance between propositions would be to solve a central problem in philosophy: the problem of providing an adequate description of what a proposition is 'about'[11]. It is not clear that any solution to that problem is forthcoming. For our present purposes, a partly syntactic definition of relevance will have to do. Are these relations too hard to compute? No harder than calculating purely logical notions of relevance (as those above) that define relevance in terms of logical relations between formulas. In our definition, to calculate the relevance relation for a pair of formulas, we need to determine the smallest language for each formula (Herzig and Rifi [55] have shown that this problem is in co-**NP**). However, purely logical notions of relevance typically involve several iterations of the derivability problem (which is of equal complexity).

An advantage of the approach suggested is that the smallest language for a particular formula can be precalculated and kept associated with the formula (to avoid recalculation). The relation of relevance used above is independent of the belief representation and therefore does not need to be recalculated with any additions to a belief state. When a formula is added to a belief state, its relations of relevance with other formulas is present in its language. Adding a new input in its smallest language to a belief state will ensure that over a period of time, the belief representation has a majority of its contents in this reduced form, thus facilitating future revision.

[11] Nelson Goodman brought out the difficulty of this problem in [43].

4.3 Hansson and Wasserman on Local Change

Hansson and Wasserman [52] develop a model for local belief revision and work with the notion of a *compartment* of a formula. Given two sets of formulae A and B. the relevant formulas in B for A are defined as those that contribute for proving or disproving any formula of A. This relies on the notion of a *kernel set*. a minimal subset that implies a given sentence.

DEFINITION 20 [50] *Let C be an inference operation on $\mathcal{L}$. Then the kernel operation $\perp_C$ is the operation such that for all $B \subseteq \mathcal{L}$ and $\alpha \in \mathcal{L}$. $X \in B\perp_C\alpha$ if and only if:*

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

The elements of $B\perp_C\alpha$ are called α-kernels ($\perp_{Cn}$ is the kernel operation for Cn. the classical consequence relation). The compartment of B around A is given by the union of the α and $\neg\alpha$ kernels of B for all elements α of A. The inconsistent kernels are simply dropped (they imply all sentences in the language).

DEFINITION 21 [52] *Let $A. B$ be sets of sentences. The A-compartment of B. is defined as: $c(A, B) = \bigcup_{\alpha\in A}(\bigcup((B\perp_{Cn}\neg\alpha) \cup (B\perp_{Cn}\alpha)\ (B\perp_{Cn}\perp))$. c is called a compartmentalization function.*

These compartments can overlap and, in distinction to the **LS** model, cannot and should not be viewed as a partition of the belief set into different topics or subjects[12].

[12]An interesting connection with the notion of relevance as developed in the sequences model is that for some sentences α, β, γ, it is possible that α is not relevant to β or γ but is relevant to $\{\beta, \gamma\}$. As an example (due to Dubois), take α = "I take a bath", β = "I use a hairdryer", γ = "I die".

OBSERVATION 3 [52] *1. For all sets* $A, B \in \mathcal{L}$, $c(A, B) = c(A. c(A. B))$.
2. If $A \subseteq A'$ *and* $B \subseteq B'$, *then* $c(A, B) \subseteq c(A', B')$.

While the definition of the compartments is couched in purely logical terms. they depend on the belief base. Hence, additions and deletions to the base change the compartments and the set around which the compartment is defined. This gives us a context for the retrieval of relevant formulas (another point of similarity with the notion of relevance developed in the sequences section). The *localization* of a given inference operation C to A is defined as follows:

DEFINITION 22 [52] *Let* C *be an inference operation on* $\mathcal{L}$ *and let* c *be a compartmentalization function. Then for any set* A, *the* A*-localization of* C *is the inference operator* C_A *such that for all sets* B *of sentences:* $C_A(B) = C(c(A. B))$. *A set* B *is* A*-locally consistent if and only if* $\bot \notin C_A(B)$.

The inference operator C_A has the following properties:

THEOREM 4.2 [52] *Let* C_A *be the* A*-localization of an inference operation* C. *Then:*
If C *satisfies monotony, then so does* C_A.
If C *satisfies monotony and compactness, then so does* C_A.

It is then shown local versions of the operations of contraction, revision, consolidation and semi-revision can be constructed and characterized by means of postulates.

Hansson and Wasserman point out that the method of constructing compartments is extremely inefficient. The representation results, however, do not depend on the way the compartments are defined, but only on the properties of the local inference operation that is defined. Wasserman [106] has provided some additional techniques for efficient retrieval of relevant information from the belief base.

Chapter 5

The $\mathcal{B}$-structures model

We now propose a new model for representing and revising belief structures, which relies on a notion of *partial language splitting* and will tolerate *some amount* of inconsistency while retaining classical logic *locally*. This model will aim to:

- Provide a representation for beliefs that is relevance sensitive and inconsistency tolerant.
- Provide psychologically plausible, computationally tractable procedures for belief revision.
- Provide a query answering scheme that preserves an agent's ability to answer queries in a coherent way from a possibly inconsistent epistemic state.
- Preserve the distinction between implicit and explicit belief.

The $\mathcal{B}$-structures model will extend the *splitting languages* framework by incorporating its intuitions and going beyond them. That framework showed, using a partly syntactic notion of relevance, how an agent's beliefs could be (uniquely) divided into sub-areas and how this division could be used in belief revision. It did not address the issue of explicit versus implicit belief, nor did it tackle the question of inconsistent beliefs. The B-structures model will do both. We will represent an agent's beliefs, not by a *theory* as is usual, but by what we call a B-structure, which generalizes the notion of a theory. Portions of this study were originally presented in [11].

5.1 Modifying the LS Model

Relevance Sensitivity

One amendment to the **LS** model that suggests itself is that subtheories contain beliefs that, while more relevant to one another than to beliefs in other subtheories, are not *irrelevant* to beliefs in other subtheories. The disjointness condition of the **LS** model should be amended to a model in which the languages of the subtheories have some overlap or crosstalk, or, set-theoretically, that the partition of the language is not a strict partition. This relaxation is psychologically realistic. Consider the following example: Bill Clinton's problems with Monica Lewinsky are connected with the bombing of Iraq, which may affect the price of oil and which in turn may affect my airfare to India. Whatever their connections, we do not reason with all these beliefs together at one time. More often than not, they will be separated, though only partially; 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.

We want a model in which beliefs about one subject matter are stored together in one subtheory and those beliefs can be relevant to beliefs about other subject matters. There might be agents that partition their beliefs exclusively: these correspond to the agents modeled by the **LS** model. Some subtheories will be irrelevant to others, as in the case of an agent's beliefs about tea and the agent's beliefs about the magnetohydrodynamic properties of plasma, but it will not be the case (in general) that our beliefs are partitioned into sets where for any pair of belief sets the two will be irrelevant to one another. If we have relevance between two sets, then we have an overlap of symbols between the two languages. However, it is not specified within this model *when* two theories are relevant to each other; we assume that if the theories share symbols, then each theory contains formulas that are relevant to formulas in the other.

What we propose is a modification of the **LS** model in which two distinct sublanguages are not necessarily disjoint: it may be the case that $L_i \cap L_j \neq \emptyset$ for distinct i, j. We want it to be the case that each of the subtheories is individually consistent, that is, we want that $\forall i, T_i \nvdash \bot$. Yet, we want to allow for the possibility that $\bigcup T_i \vdash \bot$. The **LS** model is a partially correct model in describing the partitioning of beliefs in our mind: it is incomplete in not taking into account the *limited* relevance that these sets *might* have to one another. We explain the splitting as a splitting of beliefs into exclusive zones about unrelated affairs and yet, in Quinean fashion, since there is linkage amongst our beliefs, we do not have *exclusive partitioning*. A model that builds in such a limited relevance condition in terms of a limited overlap amongst language partitions is a general model; one of which the **LS** model is a particular case.

Minimal Change

When such a belief structure is revised in the face of new information, we would like to retain the intuition that the new belief structure be as similar to the old one as possible. We will try and do justice to Quine's intuitions about minimal change to beliefs and Gärdenfors' *preservation criterion*. Minimal change is dependent on the representation for beliefs and the limited relevance model will aid in this notion.

Computational Tractability

Such an overlap does not affect the computational issues addressed by the **LS** model since our primary concern will be in ensuring that the relevance does not become *trivial*. If there were a relation of trivial relevance, every belief would be relevant to every other belief and any computational task would be intractable. We seek a middle ground between the zero relevance suggested by the **LS** model and the trivial relevance suggested by classical models of rationality. Since we seek to represent the beliefs of real people, and since the satisfiability problem is **NP**-complete, even at the propositional level we will seek to be aware of the computational limitations of real agents.

Inconsistency Tolerance

Global inconsistencies cannot be represented in the **LS** model. Suppose an agent believes theories T_i in languages L_i and the L_i are mutually disjoint. Then if the T_i are individually consistent, they are also jointly consistent. Thus the **LS** model cannot explain how an agent can be locally logically omniscient – i.e., derive logical consequences *within* each L_i but still fail to be globally consistent. Since the **LS** model requires that these languages be disjoint, there is no answer to be found on how someone could entertain inconsistent beliefs.

We would like to model an agent who is locally logically omniscient, who does not believe any outright inconsistencies, but whose global belief structure might well be inconsistent without its being aware of it. This means, that as in the demand for limited relevance, the languages L_i must be allowed to overlap on occasion. Our approach will be distinct from paraconsistent approaches in that we will allow full use of classical logic, locally, while using a multi-valued logic for answering queries.

The Explicit - Implicit Belief Distinction

Since the agent's beliefs can be inconsistent, we will need to keep the B-structure logically open and define a companion notion of inference in order to define the set of implicit beliefs. This inference operation must block the derivation of arbitrary beliefs from the agent's belief representation and avoid the charge of absurdity. We want to maintain the distinction between a finite set of explicit beliefs and those beliefs that can be derived from that set. There are limitations: logical models of belief representation are unable to capture fully the intuition that there are implicit beliefs that are not obtained by systematic deduction from a set of explicit beliefs. So long as the model captures the intuition that there is a principled distinction to be made between the primary belief representation and the beliefs derivable from it, we will take ourselves to have made some progress.

Three basic definitions are all we need to establish the B-structure model.

DEFINITION 23 *A belief structure $\mathcal{B}$ on L is a set $\{(L_1, T_1), ..., (L_n, T_n)\}$ such that $L = \bigcup L_i : i \leq n$, and each T_i is a consistent, finitely axiomatizable theory in L_i. The T_i are $Cn(\Gamma_i)$ where the Γ_i are the explicit beliefs of the agent in language L_i. If $n = 1$, then $\mathcal{B}$ will be just a theory.*

A B-structure is a collection of sets of beliefs, each one of which is individually closed and consistent. When we keep them apart, we often fail to notice that this might involve us in inconsistencies. Querying, however, a sort of Socratic prodding, can expose this under the right conditions.

The requirement of maintaining a B-structure as a set of consistent theories captures the intuitive feeling that most of us have of being 'reasonably' rational agents. In so thinking, we take ourselves to be aware of the beliefs implicit in our explicit beliefs. Yet, these tend to be just a small subset of our total set. Being asked to consider whether we are aware of our implicit beliefs normally leads us to evaluate one theory at a time. Considering larger and larger portions of our beliefs can force us to drop our earlier claims of being aware of all the consequences of our beliefs.

As in the case of belief bases, the object of revision is a set of beliefs that is not necessarily logically closed and we now get an idea of how local change can be made coherent. The definition of a B-structure above is a mixture of the foundationalist and coherentist approaches since its components, the subtheories T_i, are logical closures of sets of explicit beliefs. This enables us to both isolate a class of special beliefs and make a distinction between different inconsistent B-structures which under closure would be identical. A distinction should be made here with the notion of *mixed representations* as used by Nayak [73]. A mixed representation is the closure of more than one set and is not identical to what we have suggested here.

Since we want to retain *some* amount of disjointness, the following definition captures the notion of limited overlap:

DEFINITION 24 *Given a finite propositional language L and a set of languages $\{L_1 \ldots L_n\}$ such that $L = L_1 \cup \ldots \cup L_n$, $\{L_1 \ldots L_n\}$ is a k-partition of L if any propositional symbol q occurs in at most k of the languages L_i.*

The overlap of symbols may be thought of as "cross-talk" amongst the T_i; beliefs in some of the T_i may be relevant to beliefs in others. A k-partition with a smaller k will be *more* disjoint than one with a larger k and indicates a finer partitioning of her beliefs by the agent. In particular, a normal partition is just a 1-partition. A k-partition is *a fortiori* a $k+1$-partition and hence, for instance, a maximally fine 2-partition will usually be finer than the finest 1-partition whose existence is guaranteed by Lemma 1 (Chapter 4). Tolerating some overlap makes it easier to organize one's beliefs in smaller chunks.

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

Given the definition of relevance (see Definition 15. Chapter 4) carried over from the language splitting model. the motivation for the definitions of the B-structures model is clear. If a B-structure is in languages L_1 and L_2. and $\alpha \in T_1$ then there might also be a formula $\gamma \in T_2$ such that $L_\alpha \cap L_\gamma \neq \emptyset$: α is relevant to γ. If B consists of T_1 and T_2, and $\alpha \in T_1$, then for any $\beta \in T_1$, and any $\gamma \in T_2$. α is more relevant to β than to γ and β is more relevant to α than to γ. We place beliefs in subsets according to how relevant we take them to be to one another: this is a subjective placement, dependent on the agent. We might inspect an agent's B-structure and expect that objectively, a different splitting could have been carried out (because of the languages involved).

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

DEFINITION 25 *A belief structure $\mathcal{B}$ as in Definition 23 is m-consistent iff any m of the T_i are jointly consistent.*

We do not require that the whole collection be consistent, but the *amount* of inconsistency is limited by this requirement. If $\mathcal{B}$ is 1-consistent, that merely says that all the T_i are individually consistent. At the other extreme, if $\mathcal{B} = \{(L_1, T_1), \ldots, (L_n, T_n)\}$ is n-consistent, then it will be consistent, period. The level of m-consistency is a measure of the coherence of the agent's beliefs[1]. In terms of the models of $\mathcal{B}$, if $\mathcal{M} = \{M_i | M_i \models T_i\}$, then $\{\mathcal{M}_1 \ldots \mathcal{M}_n\}$ may have the property that the $\bigcap\{M_m : m < n\} \neq \emptyset$ but $\bigcap M_i = \emptyset$.

In between 1-consistency and n-consistency lies an entire gradient of beliefs: between the ignorant and the omniscient lies all of humanity. Most human agents might acknowledge their inconsistency in some matters and, yet, take themselves to be consistent in most of their reasoning. As we put more theories together to reason with, the greater the chance that we make our set of beliefs inconsistent. It is this aspect of reasoning that m-consistency tries to capture. With the notion of m-consistency the agent tolerates an inconsistency that does not render its belief set unusable[2].

Examples of B-structures

The Bald Person: A natural example of such a collection arises in the context of the *bald person paradox*. Suppose we say that a person with 10,000 hairs is not bald but that a person with 0 hairs *is* bald. We also say that a non-bald person cannot become bald through the removal of just one hair. This gives us the axioms $B(0)$, $B(n) \rightarrow B(n+1)$, for $n = 0$ to $n = 9,999$, and finally, $\neg B(10,000)$. These 10,002 axioms are inconsistent, but any 10,001 of them are consistent. This gives us a 10,001-consistent, inconsistent collection of axioms, all of which are accepted in practice. This example suggests that we can coherently speak of B-structures being formed with respect to particular situations and contexts; in this situation, a

[1] A similarity may (with similar motivation) be noted here with Kyburg's [61] *Principle of n-wise Consistency*: If K is a body of reasonably accepted statements, then for no set of statements $s_1, s_2, s_3 \ldots s_{n-1}$ is it the case that each of $s_1, s_2, s_3 \ldots s_{n-1}$ and $\neg(s_1, s_2, s_3 \ldots s_{n-1})$ is a member of K.

[2] We can draw a parallel here with the notion of the coherence level of a classical theory as used by Schotch and Jennings with the difference being that the agent's belief structure is not a closed theory split up into clusters, but a set of theories the union of which is not closed under classical consequence.

B-structure that contains our beliefs about the baldness predicate. This requires an extension of the definitions above that we defer for the time being.

The Wily Airline: For a more practical example, consider an airline which has 100 seats on a flight, and accepts 120 reservations. If customer number i is C_i, then the airline is asserting, for $1 \leq i \leq 120$, that C_i has a seat and also that there are only 100 seats. This is an inconsistent set of assertions, whose inconsistency will not matter if 100 or fewer of the customers remind them of the respective assertions by showing up at the gate. In response to the query "Do I have a seat?", the airline is prepared to say "Yes" 120 times[3]. I make no claims for the ubiquity of such a style of hiding inconsistency but it certainly seems prevalent.

A Four Valued Semantics

To represent how an agent *uses* its B-structure and revises it, we make use of a four valued semantics provided by multivalued logics such as those devised by Belnap [5] and Fitting [24, 26] where the four truth values are $\bot$, *true*, *false*, $\top$, which stand for, respectively, *no information*, *true*, *false*, and *over-defined* (or inconsistent). In Belnap's logic $\mathcal{FOUR}$, truth values can be thought of as *sets* of ordinary truth values. $\bot$ for $\emptyset$ (take it to indicate a lack of information), $true$ for $\{true\}$, $false$ for $\{false\}$ and $\top$ for $\{false, true\}$. In this four valued lattice there are two *partial orderings*.

The *knowledge* ordering ($<_k$) : In this ordering $\bot <_k$ *true* and *false* $<_k \top$. Generally, if α is below or the same level in the ordering as β, we can write $\alpha \leq_k \beta$. *true, false* are *incomparable* under the knowledge ordering. In the knowledge ordering, an increase means that at least one or more of the propositional atoms gains an assignment of *true* or *false* that it did not have without losing any previous truth assignments that it previously had. The knowledge ordering represents degrees of

[3] I thank Henry Kyburg for pointing this out.

information: so, obviously, having no information, $\bot$ belongs at the bottom of the ordering and, having too much information (to make up ones mind), $\top$, is at the top.

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

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

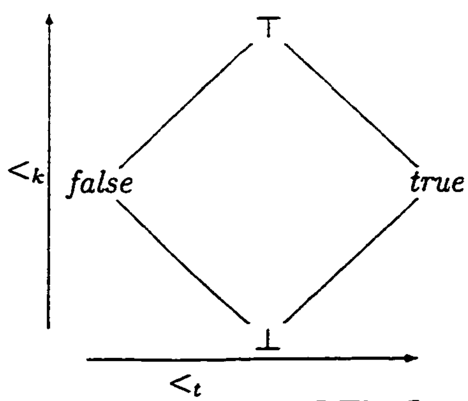

Fig. I The Logic $\mathcal{FOUR}$

Why a Four-valued Logic for Query Answering?

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

> Doing so provides a natural framework for the semantic understanding of logic programs that are *distributed over several sites, with possibly conflicting information coming from different places* [4]

[4]Fitting, [25].

Now replace the phrase 'logic program' with belief structures, and the application of such a truth value space to the answers given by a possibly undecided, partly ignorant, partially omniscient, resource bounded agent (i.e., a realistic agent) becomes apparent. The possibility of allowing dependencies between parts of an agent's belief structure that affect that agent's responses to a particular question are also hinted at in [25] where dependencies are allowed between logic programs located at different sites with the ultimate response often being a conglomerate of the conflicting answers.

In classical logic we are asked to think in terms of a dichotomies of evidence for a given proposition–totally convincing and totally unconvincing. In multi-valued logics, the idea is to drop a false dichotomy. Such dichotomies do little justice to how we often respond to questions about our beliefs. Allowing a four valued space for responses lets us make a distinction between those implicit beliefs to which we take ourselves to be fully committed, and those beliefs that we only weakly believe. We might have sets of explicit beliefs to which we take ourselves as fully committed, and we might yet shrink from some of their consequences.

5.2 Answering Queries

Our model shows that an agent whose beliefs are represented by a B-structure is able to answer questions coherently. An agent with an inconsistency present in its beliefs will not respond in the affirmative to every query: the agent will escape the damning indictment of having a vacuous belief set if the queries it is subjected to are short enough to only activate part of its belief structure. We now provide a formalization of the intuitions at hand. Given a formula α and B-structure $\mathcal{B}$ the query "α?" is answered by giving the value $v_{\mathcal{B}}(\alpha)$ – abbreviated as $v(\alpha)$ – which is defined as follows:

DEFINITION 26 *Let* $\Gamma_\alpha = \bigcup\{T_i : L_i \cap L_\alpha \neq \emptyset\}$, *where* L_α *is the smallest language in which* α *can be expressed.*

If Γ_α *is consistent, then*

if $\Gamma_\alpha \vdash \alpha$, *then* $v(\alpha) = true$

if $\Gamma_\alpha \vdash \neg\alpha$, *then* $v(\alpha) = false$, *and*

$v(\alpha) = \bot$ *otherwise.*

If Γ_α *is inconsistent, then* $v(\alpha) = \top$.

Intuitively we see *which* of the theories T_i *could be* relevant to α and put them together to get Γ_α. Γ_α is used to answer questions about α and the rest of the theories T_j are *not* brought into play. We can think of the Γ_α as being the *evidence factor* for a particular proposition. The longer the formula (i.e., the greater the number of distinct propositional symbols in the formula) the greater the chance that larger parts of the B-structure will be activated: the chance is greater that L_α will overlap many languages in the B-structure. Based on the derivation of the query from the evidence factor associated with it, the agent assigns it a ranking on the truth orderings. The agent is able to reason by pulling in information from multiple, yet relevant subject matters and is able to offer appropriate responses to situations that call for such an integration.

Example: When someone asks me whether I believe the Yankees will win the World Series this year, I do not consider information on whether the Iditarod race will again have a female winner to be relevant. I might collect all the relevant information to answer the question such as last year's records, other teams in the leagues, and so on, but that is as far as I will be willing to go. Or to return to the example of the pilot in Chapter 1, were the pilot to be asked on the radio whether any enemy aircraft had been sighted, he would respond with the understanding that the only relevant information is his sighting and that information about weather in the area is irrelevant[5].

Limited Consistency:

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

THEOREM 5.1 *Provided that L_α has at most l distinct symbols. the L_i are a k-partition, and $m \geq k \times l$, then Γ_α (as in Definition 26) will be a union of at most m of the theories T_i. If the collection $\{T_i : i \leq n\}$ is m-consistent. then Γ_α will be consistent and exactly one of the first three values will be given.*

Proof: Immediate from definitions.

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

Another way to think about the query answering of an agent that organizes its beliefs in a B-structure is to think about the agent and its questioner as engaged in a *prover-adversary* game[6]. The agent with a B-structure such that $\bigcup T_i = K_\perp$ asserts

[5]What is certainly not currently captured in this model is that I might stop at a certain depth and not consider any more information because it would require too much effort on my part. This requires a notion of more fine-grained relevance or a scheme for ranking some formulas higher than others in terms of relevance to a given proposition. For the time being, we defer such an investigation.

[6]Pudlak and Buss, [85].

that there is a truth assignment t that satisfies $\bigcup T_i$. Since there is not any such truth assignment, the prover asks the adversary the truth values under t of various formulas. The adversary loses if it assigns **T** to say $\alpha \wedge \beta$ but **F** to $\alpha \vee \beta$ or makes any other such error which is in conflict with the truth tables of the various logical connectives. Since such a truth assignment cannot exist, the agent will ultimately be caught in an inconsistency. However, if the prover is allowed only a *bounded* number of questions, the agent can get away with its concealed inconsistency. Consider the following example due to Pudlak and Buss [85]:

The Prover-Adversary Game: The formula $(\neg)^n(\alpha \wedge \neg\alpha)$ requires $\log(n)$ questions to prove the adversary's inconsistency. Suppose $n = 2^k$ and the prover asks the adversary the truth value of $(\neg)^{2^{k-1}}(\alpha \wedge \neg\alpha)$. If the agent gives the truth value of **T** (which is the wrong value) then n has been reduced by half and the prover can now work on the formula $(\neg)^{n^{2^{k-1}}}(\alpha \wedge \neg\alpha)$. If the adversary now gives the value **F** there will be an inconsistency between the truth value **F** for $(\neg)^{n^{2^{k-1}}}(\alpha \wedge \neg\alpha)$ and the truth value **T** for $(\neg)^{n^{2^k}}(\alpha \wedge \neg\alpha)$. The distance between the two formulas in terms of negation is now 2^{k-1}. Each time the prover spends one question, the distance between a correct value and an incorrect one is reduced by a factor of two. After at most $k = (\log n)$ steps, the distance will be one, and the agent will be caught.

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

DEFINITION 27 *An agent with belief structure $\mathcal{B}$ implicitly believes the proposition α iff $v(\alpha) = true$. Therefore, the set of implicit beliefs of the agent is the set of formulas which receive the valuation 'true' under the query answering scheme (as in Definition 26) $= \{\gamma | v_{\mathcal{B}}(\gamma) = true\}$*

If so, it is true that α follows from the agent's explicit beliefs but the converse does not hold in general. If the explicit beliefs are jointly inconsistent, then their consequences will be the absurd belief set, but *the agent may still have implicit beliefs which are a reasonable set.*

5.3 The Fate of Adjunction

What about the rule of adjunction in such a model? If an agent implicitly believes α and β, does it also implicitly believe $\alpha \wedge \beta$? That is, is the rule of adjunction $(\alpha, \beta \vdash \alpha \wedge \beta)$, preserved? The answer provided by the B-structure model is sensitive to the *languages* in which the formulas α and β are expressed (i.e., the subject matter of the propositions) and to the *length of the formulas*. The longer the formulas, the greater the chance that we will not reply in the affirmative to the question of whether we believe in the conjunction as well. So, if $v_{\mathcal{B}}(\alpha) = true$ and $v_{\mathcal{B}}(\beta) = true$, and $|L_\alpha| = l = |L_\beta|$ and $|L_{\alpha\wedge\beta}| \leq l$, then $v_{\mathcal{B}}(\alpha \wedge \beta)$ will, more often than not, be *true*. However, it is easy to find cases where the longer formula $\alpha \wedge \beta$ forces the agent to simultaneously consider several of its beliefs and may result in the detection of the underlying inconsistency.

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

Answering a conjunction as above has the agent assigning *true* to both conjuncts but on being asked whether the two are true together as a conjunction we cannot ascribe an unambiguous truth value to it. It seems true to us under one reading, and yet, since responding to the conjunction has forced us to deal with more information than we did before (corresponding to the activation of more theories by a longer formula) we might consider the conjunction to be false. Hence, the answer of T. We now have a *principled reason* to reject the unrestricted application of the rule of adjunction. We might not accept the conjunction of statements that are large in number since there is a high probability that forming the conjunction of those statements exposes associated inconsistencies.

What does the failure of the rule of adjunction tell us? The answer lies in our motivations for defining this sort of inference. The development of an explicit representation of beliefs and the definition of a method for revising that set of beliefs in the light of new information entails the definition of a new form of inference from that representation. Whether this corresponds to the rules of inference associated with propositional logic is an interesting question, but not one that is supposed to function as a guiding light. The failure of standard rules of inference is in the tradition of non-standard propositional logics such as relevance and four-valued logics. In each of these systems the motivational point of inquiry has been the observation that certain rules of inference were inappropriate for some epistemic contexts. In our case, we have by noticed that the rule of adjunction leads to paradoxes and that logics constructed by dropping the rule offered a way of tolerating inconsistency in a set of beliefs and defining plausible methods of inference from that set[7]. The case against adjunction is older, however, and a more radical case against it can be made.

Against Adjunction

One issue common to puzzles of beliefs such as the Lottery Paradox[8] and the Pref-

[7] We have already noted that promising approaches to modeling the holding of inconsistent beliefs often come from logics that explicitly drop the rule. I thank Henry Kyburg for pointing out that there exist systems of paraconsistent logics such as those of DaCosta [13] that do not drop the rule of adjunction.

[8] See Kyburg [60] for arguments against the notion that the lottery paradox can be solved by a

ace Paradox is the failure of the principle of adjunction. The principle (in epistemic contexts) can be expressed as the following equivalent statements:

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

If the set S above is the set of implicit beliefs that are derivable from a B-structure, then this principle fails to hold in general. Indeed, when the set above is the set of beliefs of any agent, it seems that the principle is likely to fail. A strong and simple argument against adjunction is easily made. We clearly believe that we have false beliefs. If we accept the rule of adjunction then to accept any set of beliefs is to be entitled to accept their conjunction. To believe, however, that one of our beliefs is false is to deny such a conjunction. Kyburg [62] has suggested that the rule of adjunction should be dropped as a rule of inference and has gone so far as to say that[9]:

> the unsoundness of adjunction is exactly what divides real-world inference from the ideal inference of the classical logicians: and so it should.

What makes the rule of adjunction so problematic? The presence of the Lottery and Preface paradoxes for one, but it shows up in other unproblematic areas of reasoning such as *inference to the best explanation* (this example is due to Kyburg [62]). If we have two explananda E_1 and E_2, then why should we believe that the conjunction of the best explanation of E_1 and the best explanation of E_2 will be the best explanation of $E_1 \wedge E_2$? The best explanation (given what we know) of a failure of the accounts

statistical version of the puzzle.

[9] [62], page 124.

to balance, E_1, might be that the accountant made a mistake: the best explanation of the accountant's trip to Brazil, E_2, might be that he wants a winter vacation. Given this background knowledge, the best explanation of $E_1 \wedge E_2$ might be fraud; an application of the rule of adjunction could lead us to a conclusion that is suspect.

Is it enough to consider *restricted application* of the rule of adjunction, such as that suggested by non-adjunctive logics we considered previously[10]? The Rescher and Brandom logic of inconsistency, as we have seen, concerns itself with semantic paradoxes. Kyburg considers their approach to be insufficiently opposed to the rule of adjunction. It is only in inconsistent worlds[11] that the rule of adjunction fails, and yet we can easily imagine a world in which we have 10^{10} measurements each of which is accurate within reasonable limits, and yet we would not want to form the conjunction of those statements since we would be justified in being sure that some of those are wrong. There is nothing inconsistent about such a set of beliefs, and yet it is one in which the rule of adjunction would fail as well.

Schotch and Jennings' approach considers maximal consistent subsets of a body of statements that are globally inconsistent. We can then think about statements that are in every such subset, statements that are in some such subset and so on. What is applicable to the entire corpus is a rule as follows: if the number of maximal consistent subsets of a set of statements is two, then the number of premises that can be combined is three. We can combine α, β, γ by adding the pairwise disjunctions of these statements $((\alpha \wedge \beta) \vee (\alpha \wedge \gamma) \vee (\beta \wedge \gamma))$. One of these conjuncts at least must appear among the classical deductive consequences of one of the maximal consistent

[10]We have already looked in some detail at the Rescher and Brandom logic of inconsistency [89] and the Schotch and Jennings approach [98]. Further details on the Schotch and Jennings approach are presented in [97].

[11]See the earlier discussion in Chapter 3 for a definition of this term.

subsets. In this approach, then, maximal consistent sets of subsets are treated as conjunctions.

This, for Kyburg, is too permissive. In the case of the lottery, a maximal consistent set of sentences describes the result of the lottery: *ticket number 45555 wins and all other lose.* It is not reasonable to conjoin arbitrary subsets of such a set of sentences even though they are consistent. Furthermore, a million ticket lottery gives rise to a million and one cells (the number of maximal consistent distinct subsets of our assumptions). However, if we are expanding a subset of an inconsistent body of knowledge to the point where the next statement to be accepted will produce an inconsistency, then we are justified in regarding the conjunction of the statements that came before as being infected by an inconsistency as well[12]. Kyburg takes this to indicate that all forms of adjunctive inference should be dropped. This might strike some as too restricting, but there are caveats:

> If we have reason to accept p_1 and to accept p_2 and...to accept p_n we may have good reason also to accept their conjunction $p_1 \& p_2 \& \ldots \& p_n$. (If we do not, their use as premises is surely suspect!). We need only deny that this is always the case, for any ps and any n which is what the rule of adjunction *requires* of us

Kyburg suggests a more austere position:

> ..a persuasive argument proceeds from essentially one premise: an argument from a set of premises (P_i) to a conclusion Q is acceptable *just in case there is some finite acceptable conjunctive premise* $P = P_j \& P_k \ldots P_l$ *where each conjunct is in the set* P_i *such that* $P \supset Q$ *is a theorem* .

Such a requirement clearly blocks the unrestricted application of the rule of inference for conjunctions of arbitrary length.

It certainly seems that the adjunction principle fails (in general) in inference from B-structures, though perhaps not quite in the fashion envisaged by Kyburg. The

[12] Kyburg [62], page 124.

reasons it fails ought to be of interest to someone like Kyburg as well. though, because the reason it fails regards the size of the conjunction in question and on the subject matter of the conjunction and its conjuncts. Nothing more could be asked for in this case. The reasons for its failure are good ones; we are often prepared to affirm conjuncts whose simultaneous truth is dubious; it is the putting together of those very same conjuncts that strikes us as implausible[13].

The B-structure model suggests (as Kyburg seems to as well) that a rule of *partial adjunction* be adopted. The wholesale adoption of the rule is what can get us into trouble. What the B-structure model lets us do is formalize this failure of adjunction and come up with a plausible model for belief change as well. There is a difference between the mere proximity and the actual conjunction of propositions: there are some conjunctions that we are not willing to affirm while we might be willing to affirm each one of the conjuncts and there are certain conjunctions that we are willing to affirm precisely because we assent to each of the conjuncts.

The Lottery Paradox

The Lottery and Preface paradoxes both involve long conjunctions and the agent being unwilling to assent to the conjunction of its beliefs. These situations are easily formalized by the B-structure model; we have a high value of m-consistency in the B-structure associated with our reasoning about the lottery. Given a million lottery tickets, we believe $l_1 \ldots l_n$ where l_i = *Lottery ticket i will not win* and $n \leq 1000000$. and finally, $\neg(l_1 \wedge l_1 \ldots l_{1000000})$. These 1000000 statements are jointly inconsistent.

[13]Given the arguments above, perhaps further arguments against adjunction are unnecessary. However, it is interesting to note the results in (Vardi, [102]) which argue in favor of a logic of knowledge that is not adjunctive: the ability of human beings to reason from multiple items of knowledge raises the complexity of epistemic logics from **NP** to *PSPACE*. As an epistemic strategy. it might often be a more prudent strategy to reason from unary rules of inference rather than binary rules of inference. This provides a complexity grounded reason to use a non-adjunctive logic (or a logic that does not use binary rules of inference).

but any 999999 of them are consistent. This gives us a 999999-consistent, inconsistent collection of axioms, all of which are accepted in practice. The higher the value of m-consistency, the lesser the impression of paradox. Take each of the statements of the lottery paradox to be expressed in a language L_l. Then the lottery paradox says that an agent will answer in the affirmative to the questions $\alpha_1, \ldots \alpha_n$, but will give the answer $\top$ to the conjunction $\alpha_1 \wedge \ldots \alpha_n$ since $L_{\alpha_1 \wedge \ldots \alpha_n} \geq m$ where m is the level of consistency of that agent's B-structure.

The Preface Paradox

In the Preface Paradox similarly, we have a set of beliefs, $p_1, p_2, \ldots, p_n$, which we believe individually and yet, it is also the case that we do not believe in their conjunction. Given a book with 500 pages, $p_1 \ldots p_n$ are all asserted where p_i = *This page does not contain an error* and $n \leq 500$, and finally, $\neg(p_1 \wedge p_1 \ldots p_{500})$ is also asserted. These 500 statements are jointly inconsistent, but any 499 of them are consistent. This gives us a 499-consistent, inconsistent collection of axioms, which we find reasonable to work with. It is also plausible to suggest that in the case of a not-so-large book, we would be more likely to be sure that we have not made any errors in the text. Take $p = 2$ above and we might be willing to assent to the conjunction. As the number increases however, we cease to retain such faith in our assertions. What this suggests is that we often are justified in adjoining premises when those premises are small in number. It is when they are long enough that we run into trouble. To allow a rule of inference at all times is to not pay heed to the length of the combinations thus formed; it is this disregard that inference on B-structures denies.

One objection to the use of a denial of the principle of adjunction as a way to handling the Lottery Paradox comes from views that suggest that the paradoxes are resolved by a principled distinction between *full* and *partial* belief. Such a proposal is made by Jonathan Adler [1] who suggests that the mere denial of adjunction fails to handle the paradox adequately. However, it should be noted that the approach presented here does not rest solely on a denial of the rule of adjunction. It simply says that a perfectly reasonable agent might accept the inconsistency at the heart of the Lottery paradox and that while that agent might assert (incoherently, according to Adler's view were the agent to be asserting full belief in each of the beliefs in its belief set) that each of a set of its beliefs is true including the belief that one of them is not. such an agent will, more often than not, be rational rather than incoherent in its reasoning. Our model does not make a distinction between full and partial belief: Adler's concern's are better addressed by a model that makes such distinctions.

Sorites Paradoxes

Consider the following set of axioms:

I. 0 is small.

II. For all n, if n is small, then $n+1$ is small.

III. $10^{10^{10}}$ is not small.

Parikh [79] has shown that while the system **Peano Arithmetic** with axioms I-III above is *inconsistent*, all theorems proved in it whose proofs are *short* (and are purely about numbers) are *true*. Does the B-structures model provide a similar plausible treatment of sorites type paradoxes? This method of treating sorites type paradoxes makes no claim about degrees of truth or the like. The B-structures model makes the pragmatic claim that reasoning with collections of premises taken from a belief representation such as a B-structure is not likely to implicate us in embarrassing lines of reasoning. Consider the baldness paradox case again. We have as axioms, $B(0)$, $B(n) \rightarrow B(n+1)$, for $n = 0$ to $n = 9,999$, and finally, $\neg B(10,000)$. These 10,0002 axioms are inconsistent, but any 10,001 of them are consistent. This gives us a 10,001-consistent, inconsistent collection of axioms, all of which are accepted in practice. A similar resolution is available for any other kind of sorites type paradox.

5.4 B-Structure Refinement

An agent that has explicit beliefs Γ_i in language L_i may organize these beliefs in smaller or larger subtheories. If the Γ_i are inconsistent, then we cannot organize them into a single theory[14] since that theory would be the entire belief set: some splitting is required. Organizing a B-structure into large subtheories makes computational problems harder since a more exhaustive search is required. It does have the advantage that more implicit beliefs can be derived. Organizing in smaller theories however, can mean a finer quarantining of inconsistencies and a lesser chance of being caught in inconsistency when reasoning with multiple theories. To formalize different strategies for compartmentalization of beliefs, we introduce *refinement* of B-structures:

DEFINITION 28 *A B-structure $\mathcal{B} = \{(L_1, T_1) \ldots (L_n, T_n)\}$ refines another, $\mathcal{B}' = \{(L'_1, T'_1) \ldots (L'_n, T'_n)\}$ if (i) every language L_i is a subset of some L'_j, (ii) every $L'_j = \bigcup\{L_i | L_i \subseteq L'_j\}$, and (iii) every $T'_j = Cn(\bigcup T_i | L_i \subseteq L'_j)$.*

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

Example: Suppose that a B-structure $\mathcal{B}$ has the three languages $\{p, q\}$, $\{q, r\}$, $\{r, s\}$ and theories $T_i : i \leq 3$ generated by the formulae $p \rightarrow q$, $\neg q \wedge r$ and $r \wedge s$ respectively.

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

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

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

THEOREM 5.2 *Let $\mathcal{B}$ refine $\mathcal{B}'$ and α be a formula. Then $v_{\mathcal{B}}(\alpha) <_k v'_{\mathcal{B}'}(\alpha)$.*

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

In other words, every answer that $\mathcal{B}'$ provides to a particular query will rank higher on the knowledge ordering $<_k$ than the one provided by $\mathcal{B}$. However. $\mathcal{B}'$ may give the inconsistent answer $\top$ to a query α whereas $\mathcal{B}$ may have given a *true* or *false* answer to it. Thus the answer from $\mathcal{B}'$ is more 'knowledgable' but useless! The partitioning of information into separate languages may on occasion miss some answers which might have been obtained without such partitioning, but it ensures a greater likelihood of consistency in particular queries. The more we fragment our beliefs the more likely we are to be perfectly consistent The more we merge our theories (or in the case of answering a query, the more theories we activate), the more likely we are to notice our inconsistencies. The Socratic questioner can force us through deft and persistent questioning to notice our inconsistencies. Restrict the questioning of an agent to one specialized domain and it is unlikely that any inconsistency will be noticed. Spread the net a little wider and things are not so clear any more.

5.5 B-Structure Revision

There are two strategies available to an agent for handling belief revision on belief structures and these suggest different methods for belief revision. Generally, the subtheories in which revision takes place will comprise all subtheories in which we find beliefs relevant to new information; the actual strategy followed will differ depending on context. In each case, the belief revision procedure, as in the case of the **LS** model, will be sensitive to the smallest language that the new epistemic input is expressed in. Making the revision relevance sensitive provides a principled way to select the parts of B-structure that change on receipt of new information.

Strategies for Revision

Option A: Revise all subtheories relevant to new information without merger: in this method, every theory deemed relevant to a new formula will be revised. The agent accepts the consequences *within* each sub-language while still keeping them separate. If I learn that both Beijing and London had cold winters, I am not likely to merge all my other beliefs about the two cities. The agent accepts that a new piece of information is relevant to beliefs about more than one subject matter, but decides that the reasons for keeping those subject matters distinct in the first place have not been changed radically enough by the new information to necessitate its reasoning about those areas together from that point on.

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

5.5.1 Option A - The Non-merging Option

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

DEFINITION 29 *Given a belief structure $\mathcal{B} = \{(L_1, T_1), ..., (L_n, T_n)\}$ and a new input α, for each i, the i-shadow of α is the set $\{\beta | \beta \in L_i \wedge \alpha \vdash \beta\}$. This will be a theory T_i' in L_i. T_i' is what α has to say in the language L_i. Let $\alpha_i \in L_i$ be such that $T_i' = Cn(\alpha_i)$.*

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

[15] We allow ourselves the assumption of the existence of a revision operator, taking the approach that initially, it is the choice of theories to be revised that is interesting. The choice of which part of the B-structure to revise can be thought as being *part of the choice of the revision operator* if we take the revision operator to encapsulate our epistemic strategy vis-a-vis new information. Our decision to assume the existence of an AGM (or some other) operator that works on closed belief sets is clearly a simplification.

Example: We receive the formula $\alpha \rightarrow \beta$ with theories $T_1 = Cn(\alpha), T_2 = Cn(\alpha)$. The i-shadow in this case is $\emptyset$ for both theories and no revision will be possible. The new formula has no impact on the theories of the B-structure. Here, there is a clear loss of information, since the new information has cognitive value but we do not accept it.

Example: We receive the formula $\neg\alpha \wedge \neg\beta$ with theories $T_1 = Cn(\alpha), T_2 = Cn(\beta)$. The i-shadow in this case is $\neg\alpha$ for L_1 and $\neg\beta$ for L_2.

While there is loss of information (as in the first example), this is a tradeoff that an agent accepts because of the cognitive benefits (resource wise) that accrue to the agent as a result of maintaining its beliefs in a B-structure. The following observations describe the effect of Option **A** revisions on the coherence of a B-strucure:

OBSERVATION 4 *1. k-partitions are preserved by Option* **A** *revision.*

2. m-consistency is not necessarily preserved by Option **A** *revision.*

Observation 1 is an obvious consequence of using the i-shadow for revision. Observation 2 is demonstrated below with an example. In this kind of revision therefore, we can have an agent that *adds* an inconsistency to its beliefs, thus changing the value of m-consistency while preserving the amount of overlap between subject matters. Therefore its B-structure could become progressively *less coherent* over time.

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

Examples of Option A Revision

- I switch on the television set and find out that the US Parks Service has decided to increase the number of backcountry camping permits for the next year. I revise my beliefs about the Parks Service and my beliefs about my vacation plans but there is no need to merge the two sets of beliefs. For some agents, beliefs about the Parks Service would just be part of the belief set for camping and vacations, but for an avid hiker (like me) the division of beliefs is more specialized.

- My friend Darius the musician tells me that Les Pauls are better than Stratocasters for playing blues. I believe him, he should know. I revise my beliefs about musicians who play the blues, about kinds of guitars and their qualities, about rock music as well (since I know Stratocasters are better than Les Pauls for rock songs), yet why should I merge all my beliefs about rock, blues and guitar? They are all related, yes, but why merge them all? There are blues artists who played guitar but never played rock compositions (like Robert Johnson) and there are rock guitarists who never played the blues (like Eddie Van Halen) and there are guitarists who never played either of the two (like Andre Segovia). Each one of the fields is a rich and large one. Why would I want to think about these entire sets all at once every time I thought about one of its components? They are all related, and that is why the sub-theories corresponding to them overlap, but there is no good reason for me to merge them.

- The pilot in the example in Chapter 1, on receipt of new information that the aircraft in front of him is a friendly craft, will only revise that part of his *B*-structure that stores current beliefs about aircraft sightings. There is no reason for him to revise his beliefs about the seating capacity of the aircraft or his beliefs about the presence of other enemy aircraft in the area.

Relevance Relations in Option A Revision

An alternative way to understand option **A** revision is by the notion of a relevance relation between a new input and a theory in a B-structure. The existence of such a relation is an indicator of whether the new formula brings about an increase in the information associated with a theory. In the case of Option **A** revision some theories are left untouched since they are considered irrelevant and amongst the theories that are revised, the only ones that actually change are those to whom the new formula brought new information. We offer the following definition:

DEFINITION 30 *A formula* α *is relevant to a subtheory* T_i $\mathcal{R}(T_i, \alpha)$ *iff:*

1. $L_i \cap L_\alpha \neq \emptyset$.

and:

2. $((T_i \dot{+} \alpha) \cap L_i)$ *is a proper extension of* T_i *i.e.,* $\mathcal{M}((T_i \dot{+} \alpha) \cap L_i) \subset \mathcal{M}(T_i)$.

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

OBSERVATION 5 *If* $L_\alpha \cap L_j \neq \phi$ *and if* $\mathcal{R}(T_i, \alpha)$ *then* $T_i * \alpha \neq T_i$.

That is, a new piece of information is relevant to a sub-theory if the addition of the new belief to the sub-theory results in a *more precise expression of the theory*: a new piece of information is relevant or useful, if it leads to greater specificity in our beliefs.

OBSERVATION 6 *1. If* $\mathcal{R}(T, \alpha_1 \wedge \alpha_2)$ *then either* $\mathcal{R}(T, \alpha_1)$ *or* $\mathcal{R}(T, \alpha_2)$.

2. If $\mathcal{R}(T, \alpha), \beta \vdash \alpha$ *and* $\beta \nvdash \bot$ *then* $\mathcal{R}(T, \beta)$.

3. Let $\mathcal{M}((T_i \dot{+} \beta) \cap L_1) \subset \mathcal{M}((T_i + \alpha) \cap L_1)$, *then if* $\mathcal{R}(T, \alpha)$, *then* $\mathcal{R}(T, \beta)$.

The last observation above lets us compare the relative relevance of propositions with respect to a subtheory: β is at least as relevant to a subtheory T_i as α if $\mathcal{M}((T_i \dot{+} \beta) \cap L_i) \subset \mathcal{M}((T_i + \alpha) \cap L_i)$.

The relation $\mathcal{R}$ is illustrated by the following example. Take $T_1 = Cn(p \vee q)$. $\gamma_1 = \neg(q \vee r)$, $\gamma_2 = q \vee r$, $L_1 = (p, q)$ and $L_2 = (q, r)$ Here, $\mathcal{R}(T_1\gamma_1)$ obtains, but, as may be verified, $\mathcal{R}(T_1\gamma_2)$ does not. An agent with the theory T_1 will not change its beliefs on receiving γ_2 since $T_1 * \gamma_2 = T_1$, but will do so on receiving γ_1. Informally, if I believe that (*the door is open*) **or** (*the heat is not working*) as a theory about why the room is cold, it does me some use to be told that it's not true that (*the heat is not working*) **and** (*New York won the World Series*). It is no use to be told that (*the heat is not working*) **or** (*New York won the World Series*).

5.5.2 Option B - The Merging Option

This option generalizes the procedure of [80]. Given a formula α as input and a belief structure $\mathcal{B}$ we carry out Option **B** revision by a merger of subtheories that are affected by new information. Let $\Gamma_\alpha = \bigcup\{T_i : L_i \cap L_\alpha \neq \emptyset\}$, where L_α is the smallest language of α. Let $T_\alpha = Cn(\Gamma_\alpha)$. Now replace all languages L_i such that $L_i \cap L_\alpha \neq \emptyset$ by the single language $\bigcup\{L_i : L_i \cap L_\alpha \neq \emptyset\}$, which is their union. At the same time, replace all the corresponding T_i by the theory $T_\alpha * \alpha$. This option will specialize to the procedure in [80] where the languages were all assumed to be disjoint but the receipt of information resulted in joining those theories whose languages overlapped the language of the new information. Option **B** style revision might not be the most plausible way to model revision as a general method since every new belief received leads to a merger of sub-theories (leading in turn to the progressive 'lumping' together of a B-structure over time). It is, however, a plausible method in some contexts.

Generally, Option **B** results in two formally distinct subject areas merging as the result of some new information which straddles them (in real life this is likely to happen only occasionally).

Examples:

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

- I am a graduate student starting work on my dissertation: I maintain separate beliefs about the field of belief revision and non-monotonic reasoning. I have had some exposure to both fields but have approached them relatively independently. While I realize the two are relevant to each other in that they concern themselves with formalizing plausible reasoning, I maintain separate, though interconnected beliefs about the two and do not think about the two simultaneously. I then read an article by David Makinson that points out the explicit connection and intertranslatability between the two fields. From now on, I maintain my beliefs about the two fields together. Whenever I write a paper on belief revision, I think about the non-monotonic inference operation involved, whenever I write a paper on non-monotonic logic I think about the belief revision process being modeled.

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

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

In the case of Option **B** style revision, however, it is easy to imagine a situation in which the two subtheories corresponding to bus and train timetables could be merged into a new theory. For example, the agent receives a new piece of information that connects the two: public transportation timetables for the city that the agent lives in. While compartmentalization is present in the B-structure, it is not so as to preclude the combination of different compartments for use in inference. While information exists in separate sub-theories for the agent, it is always available for query answering in contexts that are relevant to it: compartmentalization in this model does not preclude periodic mergers.

5.5.3 Alternative Revision Schemes

Other alternative revision schemes for B-structures can be devised whose full development we defer for the time being. For example, Option **C** corresponds to a slightly more specialized situation and one in which either of the options above could be followed after a preliminary adjustment of the B-structure.

Option C: In this revision scheme, an agent on receipt of information that makes an assertion of identity between two subject matters, decides to either replace two previously distinct subject matters by a third one or replaces a subject matter by another.

[16] [101], page 35.

The situation described here is actually better approximated by richer languages than we are allowing ourselves at the moment. In order to accurately model this kind of revision we need a language that contains *names* as well as *modal operators*. The need for names is obvious since the most plausible example of this kind of revision will occur in situations when we find out that two individuals that we took to be distinct are the same. So, we will need to express the formula $(I = J)$ where I, J are names. The need for modalities arises since we need to be able to express the formula $\Box(I = J)$ in order to accurately the assertion of identity between I, J.

Examples:

- I am a Babylonian sailor and navigator. In the night I navigate by the Evening Star and in the mornings by the Morning star. I think about the two astronomic entities in mutually exclusive fashion. Yet, there comes a time when I receive information that persistently connects the two and finally I learn that the two are one and the same heavenly body, Venus. At this time, I drop my decision to think of the two as separate entities, treat them as the same, drop beliefs that treated them as distinct entities and merge my two belief sets about them. I find out that instead of two stars called the Morning Star and the Evening Star, there is just one star. I could pick either of the names and use it with no fear of inaccuracy in referring. From now on, when I think about the Morning Star, I'm just thinking about the Evening Star.

- In the movie *Superman II*, Lois Lane finds out that Clark Kent really is Superman. She clearly had to drop some of her old beliefs about the two including thinking of the two as separate people and come up with some reassessments of her old beliefs, including for example, the fact that Clark is a sissy and so on. At this stage, Lois probably just started thinking of the two as alternative ways of thinking about the same person, her sweetheart[17].

Once we have carried out language (subject matter) replacement, it will be obvious that the second stage in this revision procedure will be a Option **A** or Option **B** revision. That is, on an assertion of identity, we could merge theories if we felt that the new knowledge made our treatment of the two as separate theories as superfluous. If such a need was not felt, the agent could stick to an Option **A** style revision and revise each theory separately.

First Order B-structures

The issue of first order B-structures requires considerable investigation since the introduction of a language with quantifiers introduces complications. For the time being, we content ourselves with a few preliminary remarks:

- The *i-shadow* in Option **A** will need more structure when we consider higher order languages since the definition of the i-shadow requires that each of the theories be finitely axiomatizable. This might be the case when we are dealing with a finite propositional language, but when we consider first-order languages (or others) we might not be able to make such an assumption[18]. If the theory however is axiomatizable, it is still axiomatizable by some countably infinite, recursively enumerable set of sentences. So, using revision operators for a possibly infinite set of sentences could be a technical solution for this problem.

- For B-structures, the use of first order languages with identity is not a problem the way it is in the case of the **LS** model. Theories without overlap in languages

[17]In the movie Lois Lane is made to forget at the end that Superman and Clark Kent are one and the same person i.e., the old languages are restored in her B-structure.

[18]This problem has been pointed out by Sam Buss, Rohit Parikh and Graham Priest.

can be in conflict since we are allowing inconsistency in belief structures. Unfortunately, the test of language overlap as a screen for relevance fails. We cannot deem a new belief as irrelevant to some theory T_i if there is no language overlap between the two since the new belief could conflict with T_i even if $L_\alpha \cap L_i = \emptyset$. That is, the notion of subject matter of a proposition being the language of the formula cannot be extended to first-order languages. Take for example, the theory that asserts that there are exactly two objects and the theory that asserts that there is precisely one object. The two could be in conflict without there being any overlap in symbols.

5.6 Rules of Inference for B-structures

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

- **Disjunction Introduction:**

$$\frac{\Gamma \vdash \alpha}{\Gamma \vdash \alpha \vee \beta}$$

- **Deduction:**

$$\frac{\Gamma \cup \{\alpha\} \vdash \beta}{\Gamma \vdash \alpha \rightarrow \beta}$$

- **Conjunction Elimination:**

$$\frac{\Gamma \vdash \alpha \wedge \beta}{\Gamma \vdash \alpha, \Gamma \vdash \beta}$$

- **Disjunctive Syllogism:**

$$\frac{\Gamma \vdash \alpha \vee \beta, \Gamma \cup \{\alpha\} \vdash \gamma, \Gamma \cup \{\beta\} \vdash \gamma}{\Gamma \vdash \gamma}$$

- **Modus Ponens:**

$$\frac{\Gamma \vdash \alpha \rightarrow \beta, \Gamma \vdash \alpha}{\Gamma \vdash \beta}$$

Translations of these rules in terms of valuations in $\mathcal{FOUR}$ assigned by the query answering procedure can be provided. We define inference (i.e.. we define a new inference operator $\vdash_{\mathcal{B}}$) from B-structures as follows:

DEFINITION 31 *$B \vdash_{\mathcal{B}} \gamma$ iff $v_{\mathcal{B}}(\gamma) = true$ and $\Gamma_\gamma \not\vdash \bot$ where Γ_γ is as defined in the query answering procedure (Definition 26).*

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

- **Disjunction Introduction:** If $v(\alpha) = true$ then is $v(\alpha \vee \beta) = true$?
- **Deduction:** If $v_{\mathcal{B}*\alpha}(\beta) = true$ then is $v_{\mathcal{B}}(\alpha \rightarrow \beta) = true$?
- **Conjunction Elimination:** If $v_{\mathcal{B}}(\alpha \wedge \beta) = true$ then is $v_{\mathcal{B}}(\alpha) = true$? Is $v_{\mathcal{B}}(\beta) = true$?
- **Disjunctive Syllogism:** If $v_{\mathcal{B}}(\alpha \vee \beta) = true$, $v_{\mathcal{B}*\alpha}(\gamma) = true$, $v_{\mathcal{B}*\beta}(\gamma) = true$. then is $v_{\mathcal{B}}(\gamma) = true$?
- **Modus Ponens:** If $v_{\mathcal{B}}(\alpha \rightarrow \beta) = true$ and $v_{\mathcal{B}}(\alpha) = true$ then is $v_{\mathcal{B}}(\beta) = true$?

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

THEOREM 5.3 *The inference operator* $\vdash_{\mathcal{B}}$ *for a B-structure* $\mathcal{B}$, *and for all* α, β, γ *such that* $L_\alpha = L_\beta = L_\gamma$ *where* γ *is the compound formula formed from* α, β *and logical connectives (if* $L_\gamma = \emptyset$ *then it simply consists of propositional constants true. false) satisfies the rules above.*

Proof: Immediate from definitions.

This result shows that classical rules for inference obtain in inference from B-structures when the agent restricts itself to a particular subject matter; once subject matters overlap, classical rules for inference can fail due to the interaction of beliefs about different subject matters possibly conflicting with one another (as in our example in the section on adjunction). Conversely, the plausibility of the method of query answering and inference defined on B-structures is brought out in cases where an agent might not have enough information to answer a particular query since only a small part of its B-structure has been activated. Once more information is supplied by the questioner, the agent can provide definite answers. For example, I meet a friend after a long time who enquires about a (supposedly common) acquaintance. Dimitriakos: "Does Dimitriakos have an elder sister?". I reply that I do not know of any Dimitriakos or his sister. My friend prods me "Surely, you remember Dimitriakos of Megara?". Now, I know who is being talked about. Of course, I remember Dimitriakos, we had met on my trip to Greece a year ago. I do not remember Dimitriakos (or anything about him) in isolation but I do remember him in conjunction with the information that he is from Megara.

5.7 Computational Complexity

In [41], Ginsberg echoes the paradigm definition of commonsense reasoning (at least in the artificial intelligence community) as being the process of using polynomial techniques to convert inference problems to a modified **NP**-hard problem on which exhaustive search is feasible. Our methods correspond exactly to such a technique: the complexity of a problem is a function of its size and in the case of B-structure revision, it is this input size that is reduced.

We summarize the computational costs of B-structure query answering and revision below. The complexity of the query answering and the revision procedures includes the complexity of the implication problem. The calculation of the smallest language for all new inputs is a **coNP** problem[19].

The algorithm for query answering is as follows:

- Determine smallest language of new formula (**coNP**).
- Determine theories with overlap.
- Form Γ_α.
- Test implication (**NP**).

The complexity of the revision procedures will include the computational cost of the revision for each subtheory. The algorithm for Option **A** revision is as follows:

- Determine smallest language of new formula (**coNP**).
- Determine theories with overlap.

[19] Herzig and Rifi, [55]. One often has to 'divide hard to conquer well' (I heard this from Benjamin Grosof).

- Form i-shadow (**NP**).
- Revise all relevant theories (**NP**).

The algorithm for Option **B** revision is as follows:

- Determine smallest language of new formula (**coNP**).
- Determine theories with overlap.
- Form Γ_α.
- Revise Γ_α (**NP**).

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

5.8 Properties of B-structure Revision

We present some properties that describe B-structure revision. Let $\mathcal{B} * \alpha$ denote the revision of B-structure $\mathcal{B}$ by the formula α according to option **B**. For option **A**, we need (as before in determining rules of inference from B-structures) the caveat that $\alpha \in L_i$ for some i or, at least that $\alpha \Leftrightarrow \beta_1 \wedge ... \wedge \beta_p$ where each β_k is in some L_i and p is small. However *disjunctive* information which straddles two of the L_i may be lost if we insist, as we do in Option **A**, on keeping the L_i separate. The set theoretic notions $\in, \subseteq$ are now replaced by more sophisticated generalizations which enter in with Belnap's four truth values; read $\alpha \in \mathcal{B}$ below as 'α is an implicit belief according to $\mathcal{B}$':

OBSERVATION 7 *1. $\mathcal{B} * \alpha$, the revision of $\mathcal{B}$ by α, is a belief structure.*

2. $\alpha \in \mathcal{B} * \alpha$

3. If $\alpha \Leftrightarrow \beta$, *then* $\mathcal{B} * \alpha = \mathcal{B} * \beta$

4. $\mathcal{B} * \alpha \subseteq \mathcal{B} \dot{+} \alpha$, *i.e., if* $\mathcal{B} * \alpha$ *and* $\mathcal{B} \dot{+} \alpha$ *give values* v, v' *to* α, *then* $v \leq_k v'$.

5. If α *is consistent with* $\mathcal{B}$, *i.e., it is not the case that* $v_{\mathcal{B}}(\alpha) \in \{false, \top\}$, *then* $\mathcal{B} * \alpha = \mathcal{B} \dot{+} \alpha$.

The first observation corresponds to the principle of categorial matching: after revision, we are still working with a belief structure. The second observation (which corresponds to the Success axiom) says that a new belief is contained in the revised B-structure. The third says that equivalence of revisions by equivalent formulas is obtained in B-structure revision (since each formula is expressed in its smallest language, this follows immediately). The fourth observation states that the expansion of a B-structure by a formula will always give us *more knowledgeable* answers than the revision of a B-structure by the same formula. The fifth observation says that if the new information is consistent with the B-structure, then the result of expansion is the same as that of revision.

Further Properties:

The following observations describe revisions on sub-theories T_i in a B-structure:

OBSERVATION 8 *1. If* $\alpha \not\Leftrightarrow \beta$, *then* $T_i*\alpha*\beta$ *will not always be the same as* $T_i*\beta*\alpha$*;*

2. For a sub-theory T_i *and formulas* α, β *such that* $\mathcal{M}((T_i \dot{+} \beta) \cap L_1) \subset \mathcal{M}((T_i + \alpha) \cap L_i) \subset \mathcal{M}(T_i)$, *we have* $T_i * \alpha * \beta = T_i * \beta$.

The first observation says that the order of receipt of information can make a difference in the way we revise our beliefs. The second says that the result of revising by two pieces of information, one more specific than the other is equivalent to revising by the more specific piece of information (this corresponds to Darwiche and Pearl's axiom **C1**). As an aside, note that the update procedures need not preserve refinements. If $\mathcal{B}$ refines $\mathcal{B}'$ and there is new information α, it may reveal an inconsistency in $\mathcal{B}'$, even though $\mathcal{B}$ is not so affected. This means that a belief structure $\mathcal{B}$ that used to refine another $\mathcal{B}'$ might fail to do so after revision.

Conclusion

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

In this model, minimal change is taken care of by its relevance sensitivity. A new epistemic input can only conflict with those parts of the B-structure that contain beliefs about the same (or related) subject matters. This protects the remaining parts of the B-structure from being affected by new inconsistencies. There is no notion of a trivial update at the global level in a B-structure even though such changes might take place at the local level i.e., individual T_i in a B-structure could still be subject to the trivial update operation. This seems plausible, since one can imagine scenarios in which all of an agent's beliefs about a *particular* subject matter can be displaced.

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

The issue of refinement raises a related issue, that of periodic splitting of B-structures given new contexts i.e., to devise a re-splitting operator, one that corresponds to an agent settling upon a particular, new, re-organization of its beliefs that suits its new cognitive objectives. Such a resplitting could also be undertaken by the agent after a review of its beliefs that are more exhaustive than those it normally permits itself. This could correspond to a second-order check on the accuracy and truth of its beliefs. The result of such an investigation could be a re-organization of its belief structures[20]. The formalization of such a notion must wait for the time

[20] The suggestion that reasoning agents will carry out such periodic introspection so as to ensure a certain quality of truthfullness and accuracy in their beliefs is due to Jonathan Adler.

being.

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

Chapter 6

The Belief Sequences Model

6.1 An Alternative Approach

In this chapter, we present an alternative relevance sensitive method for representing and revising sets of beliefs that are possibly inconsistent. This model corresponds to the *direct* or *vertical* approach to belief revision and tackles the issues of temporal ordering and iterated belief revision that we have had occasion to only briefly consider thus far in this study. A brief reminder of the issues: given a theory K and a proposition α, the AGM approach proposes postulates for $K * \alpha$, the revised theory with α. However, AGM-like postulates do not specify how we *came* to believe K and, after revision, it is assumed that (K and) $K * \alpha$ is a generic theory. In practice, we know that α was our *last* information. AGM postulates do not say anything about any form of reliance on the theory that is being revised, the first argument of the revision operation. This leads us to the problem of iterated revision. We have already considered some facets of this problem in Chapter 1. I now turn to a model that tackles these issues while retaining the desirable features of the approach followed thus far. Portions of this study were originally presented in [10].

We suggest that $K * \alpha$ is should not be regarded as a theory, but the *belief sequence*

$K; \alpha$ i.e., a sequence of propositions with α being the most recent. This sequence will form a kind of belief base in that it will serve as the set upon which we define an inference method to determine the agent's epistemic state. This method is inspired by our original motivations for the **LS** and B-structures model and also by work done on *ordering based representations of rational inference* by Georgatos [38], the *direct revision* approach of Brewka [8], the *ordered theory presentations* of Ryan [95] and the belief sequences model of Lehmann [63]. In [38], Georgatos shows that taking the linear order of a belief sequence as a prioritization generates a variety of inference relations. It is shown that all such schemes are non-monotonic and are therefore. compatible with belief revision. The **LS** and B-structures models have already shown us ways to ensure a relevance or context sensitive. *localized* notion of belief revision. The B-structure model also explicated the distinction between implicit and explicit beliefs by considering sets of theories which are individually consistent but can be *jointly* inconsistent: real agents often reason with an inconsistent, yet usable. set of beliefs. We want to retain all these intuitions in this model.

In a similar spirit, the method of inference of presented here will block the derivation of *explicitly inconsistent beliefs* from a *possibly inconsistent belief sequence* by using a notion of inference from *maxiconsistent subsets of relevant formulas.* Choosing maxiconsistent subsequences in order to avoid inconsistency has a long and distinguished tradition, going back to Reshcher [88]. Relevance will be determined, as in the **LS** and B-structures models by language overlap (with some context sensitivity built in).

The formula whose inference from the sequence is to be determined will impose a prioritization on the formulas in the sequence by virtue of its relevance relations with them (thus reorganizing the temporal ordering present in the sequence). Therefore. we do not treat a belief sequence as a set but rather as a linear order much like an entrenchment (see [36], [39]). Since we take relevance into account we cannot use an entrenchment ordering in our method; the relevance relations affect any orderings already present amongst the formulas.

6.2 Related Work

6.2.1 Lehmann on Sequences

The notion that a sequence of formulae captures the importance of temporal ordering and revision histories is noted and formalized by Lehmann [63]. In Lehmann's sequences model, a belief state results from a sequence of revisions—the individual revisions themselves are just those by consistent formulas. A concatenation of these sequences is denoted by $\circ$ and sequences of length one are denoted by formulas[1]. For example, we can think of the sequence $\sigma \circ \alpha \circ \beta \circ \tau$ as denoting, first, the formulas of the sequence σ, then the formula α, then β and then the formulas of the sequence τ. The belief set that results from a sequence σ of individual revisions is denoted $[\sigma]$: $[]$ denotes a revision procedure. The belief set $[\sigma \circ \alpha]$ denotes the result of revising $[\sigma]$ by the formula α. We do not however, identify $[\sigma \circ \alpha]$ with $[\sigma] * \alpha$ ($*$ is an AGM operator). The AGM postulates require that if $[\sigma] = [\tau]$, then $[\sigma] * \alpha = [\tau] * \alpha$. but in this framework $[\sigma] = [\tau]$ does not imply $[\tau \circ \alpha] = [\sigma \circ \alpha]$. Lehmann's framework allows a reasoning agent to base its revision *not just on the belief set* $[\sigma]$ *but also on the sequence itself i.e., on the history of the agent*. In this way, the temporal ordering of the beliefs plays a part in the revisions that the agent carries out. Given the

[1]We have modified Lehmann's original notation for greater readability.

definitions above, a set of axioms is presented.

Lehmann's Axioms:

(I1) $[\sigma]$ is a consistent theory

(I2) $\alpha \in [\sigma \circ \alpha]$

(I3) If $\beta \in [\sigma \circ \alpha]$, then $\alpha \rightarrow \beta \in [\sigma]$

(I4) If $\alpha \in [\sigma]$, then $[\sigma \circ \tau] = [\sigma \circ \alpha \circ \tau]$

(I5) If $\beta \models \alpha$, then $[\sigma \circ \alpha \circ \beta \circ \tau] = [\sigma \circ \beta \circ \tau]$

(I6) If $\neg\beta \notin [\sigma \circ \alpha]$, then $[\sigma \circ \alpha \circ \beta \circ \tau] = [\sigma \circ \alpha \circ \alpha \wedge \beta \circ \tau]$

(I7) $[\sigma \circ \neg\beta \circ \beta] \subseteq Cn([\sigma], \beta)$

I1 simply requires belief sets to be consistent theories and corresponds to the AGM Closure postulate. **I2** is an exact translation of the Success postulate but is stronger in that it *only* allows revision by consistent formulae. **I3** expresses the intuition that the agent should be *aware* of the result of previous revisions: if the agent were to accept β after a revision by α it should accept (now) that if α is true then β is true. **I4** says that some revisions can be made *superfluous* and can be eliminated by previous revisions. Revision by a formula that is already contained in the current belief set has no effect. **I5** deals with the case of revisions that are made superfluous by the revisions that follow. If an agent were to revise first by α and then by $\alpha \wedge \beta$, the second revision confirms all the information contained in the previous one and hence the first one is superfluous. So, if an agent were to receive information that first indicates that α is true and then later that β is also true, it should not make a difference for the agent to revise at the end by $\alpha \wedge \beta$ or to revise immediately by α and then revise later by $\alpha \wedge \beta$. Lehmann points out that this is not equivalent to revising first by α and then by β since, in revising by the conjunction, the agent learns that the two formulas are true together, while learning β alone, even after learning α could cause the agent to drop its belief in α.

I6 says that revision first by α and then β is equivalent to revision first by α and then by $\alpha \wedge \beta$. If we are given **I5**, then **I6** is equivalent to $I6'$:

$$(I6')\text{ If } \neg\beta \notin [\sigma \circ \alpha]\text{, then } [\sigma \circ \alpha \circ \beta \circ \tau] = [\sigma \circ \alpha \wedge \beta \circ \tau]$$

This postulate says that under certain conditions (i.e., if $\neg\beta \notin [\sigma \circ \alpha]$) two successive revisions can be collapsed into one revision by the conjunction of the two formulas concerned. If the revision concerned is a severe one (i.e., if $\neg\beta \in [\sigma \circ \alpha]$) then there may be a difference in revising first by α and then by β and in revising by $\alpha \wedge \beta$.

I7 is an expression of the requirement of Minimal Change. If an agent after a sequence $[\sigma]$ of revisions learns $\neg\beta$, and then β, its belief state should reflect this fact. While it believes β any additional beliefs it has can only come from the beliefs it held after the sequence σ: none of the beliefs activated by learning $\neg\beta$ should remain after learning β. Lehmann's system is weaker in one respect than the AGM system (i.e.. the non-equivalence of revisions by logically equivalent formulas) but is stronger in terms of the two postulates **I5, I7**.

6.2.2 Ryan on Ordered Theory Presentations

In [95] Mark Ryan presents a model for revision of finite sets of sentences $\Gamma = [\phi_1, \ldots, \phi_n]$, under a linear order; ϕ_n is the most recent formula in Γ. These sets (or sequences) are termed *ordered theory presentations.* Revision is carried out by appending new inputs to the theory presentation and therefore, iterated revision is handled easily. Take $\Gamma = [p \wedge q, \neg p]$ and revise by $\neg q$. The new presentation $\Gamma * \neg q = [p \wedge q, \neg p, \neg q]$.

Associated with each theory presentation Γ is its *extension*, which is the deductively closed theory that Γ presents. An extension is formed by collecting formulas starting from the most recent formulas and by giving those higher priority in cases of conflicts with formulas which occur later in Γ i.e., the construction of the extension pays explicit attention to the temporal ordering of elements of Γ. For the example above, its extension before revision is $Cn(\neg p \wedge q)$, after revision, it is $Cn(\neg p \wedge \neg q)$. Two presentations Γ_1, Γ_2 with the same elements in different order can have different extensions and two presentations which would be equivalent under classical closure can have different extensions.

Ryan then shows how to obtain models for the extensions of ordered theory presentations (it is shown that every ordered theory presentation has a model) and using those, obtains a set of axioms for revision of presentations. These are similar to those of AGM with the notable exception of axioms (K^* 4) and (K^* 8). Counterexamples to argue for the absence of these axioms can be found in [95]. Of importance to us is the use of temporal orderings to maintain revision histories, the straightforward handling of iterated revision and a construction for extensions that enables distinctions between different, inconsistent belief representations.

6.3 Belief Sequences and Relevance Relations

We begin with a preliminary definition:

DEFINITION 32 *A belief sequence is a sequence of formulae under a temporal ordering, i.e., a sequence of formulae, $\sigma = \beta_1 \ldots \beta_n$ where for any pair of beliefs β_i, β_j if $i < j$, β_j is more recent than β_i. The sequence is totally ordered: for any distinct β, α, β is more recent or later than α. Given two sequences, σ_1, σ_2 we say that $\sigma_1 \sqsubseteq \sigma_2$ if*

σ_2 is obtained from σ_1 by the concatenation of zero or more formulas ($\sqsubseteq$ is a reflexive relation); σ_1 will be referred to as an initial segment of σ_2.

Under a temporal ordering the most recent formulae occur at the tail of the sequence. We assume that each formula β_i in a sequence σ is expressed in its smallest language L_{β_i}.

We now turn to relevance relations amongst formulas in a sequence. Wasserman [106] has proposed that inference from belief bases is made more plausible by assuming the presence of some structuring relation amongst beliefs in a base. This relation facilitates the choice of beliefs to be revised, contracted or used for inference in any revision operation. Such a relation was proposed and used in the **LS** and B-structures models. We remind the reader that the language $L(\alpha)$ is the set of propositional variables in a formula α; the language L_α of α is the smallest set of propositional variables which can be used to express β, a formula logically equivalent to α. These languages were used to define the relations $\mathcal{R}_s(\alpha, \beta), \mathcal{R}_l(\alpha, \beta)$ in Chapter 4 (Definition 15). We now develop a *context-sensitive* measure for *relevance* amongst formulas in a belief sequence.

DEFINITION 33 *Two formulae $\beta_1, \beta_2 \in \mathcal{L}$ are syntactically disjoint iff $L(\alpha) \cap L(\beta) = \emptyset$. Two formulae $\beta_1, \beta_2 \in \mathcal{L}$ are logically disjoint iff $L_\alpha \cap L_\beta = \emptyset$.*

If two logically disjoint formulae are individually consistent, then they are jointly consistent. We use the word 'disjointness' here instead of 'irrelevance' since, as we see below, two formulas can be relevant even if their languages are disjoint.

DEFINITION 34 *A pair of formulas, α, β are directly relevant (or 0-relevant) if they are not logically disjoint, that is, if $L_\alpha \cap L_\beta \neq \emptyset$. Given a belief sequence σ, a pair of* *formulas α, β are k-relevant w.r.t σ iff $\exists \chi_1, \chi_2, \ldots \chi_k \in \sigma$ such that:*

i) α, χ_1 are directly relevant

ii) χ_i, χ_{i+1} are directly relevant for $i = 1, \ldots k-1$

iii) χ_k, β are directly relevant. (the sequence σ will be omitted when clear from the context).

A pair of formulas are irrelevant if they are not k-relevant for any k. Let $rel(\alpha, \beta, \sigma)$ be the lowest k such that α, β are k-relevant wrt σ (we let it be ∞ if α, β are irrelevant). Note that $rel(\alpha, \beta, \sigma)$ defined this way is a function. We write $\mathcal{R}_k(\alpha, \beta, \sigma)$ to indicate that α, β are k-relevant w.r.t σ.

Such a relevance relation is ternary as opposed to standard definitions which make it a binary relation: two formulas are relevant to each other if we can construct a connection between them by other formulas (present in the sequence) that are relevant to each of the formulas in question. The above definition extends the definition of relevance used in [80] and makes explicit the contextual nature of the relevance definition: two formulas α, β may have different degrees of relevance with respect to different belief sequences. A belief sequence defines a particular context or set of subject matters; pairs of formulas acquire different relationships to one another given differing contexts.

Example: We can think of a linkage running from the subject *Geography* to the subject matter *European Geography*, from there to *European History* and then to *History* in general, to *Military History*. While *Military History* and *Geography* are not directly relevant to one another, there are many contexts in which the two are related to one another (e.g. the military campaign in North Africa during the Second World War).

OBSERVATION 9 *If a pair of formulas are k-relevant, then,* $\forall m > k$, *they are m-relevant as well.*

This captures our intuition that if a pair of propositions are relevant at a particular level of relevance then they are relevant to each other at lower levels.

OBSERVATION 10 *Let* $[\![\sigma]\!] = \{\alpha | \alpha \text{ occurs in } \sigma\}$. *Then,* $\sigma_1 \subseteq^s \sigma_2$ *iff* $[\![\sigma_1]\!] \subseteq [\![\sigma_2]\!]$. *If* $\sigma_1 \subseteq^s \sigma_2$, *then* $rel(\alpha, \beta, \sigma_2) \leq rel(\alpha, \beta, \sigma_1)$.

OBSERVATION 11 *The relation* $\mathcal{R}_k(\alpha, \beta, \sigma)$ *is both symmetric and reflexive in the first two arguments but obviously not transitive. If* $k = 0$ *then the sequence* σ *is irrelevant to the question of relevance of formulas* α, β; *if* $k \neq 0$, *then* σ *is a parameter.*

6.4 Revision and Prioritized Inference on Belief Sequences

Revision

Revision on belief sequences is achieved by *concatenating* the new formula γ to a sequence σ. It becomes the most recent piece of information received. That is, the sequence $\beta_1, \ldots, \beta_n$ simply acquires a new member $\beta_{n+1} = \gamma$ added on at the end of the sequence.

Example: Consider the sequence $\sigma = [p, q, p \wedge q, \neg r]$. The new information $\beta = \neg p$ is appended to the sequence to give us $\sigma * \beta = [p, q, p \wedge q, \neg r, \neg p]$.

$[\![\sigma]\!]$ above is now inconsistent, and we need a notion of inference which renders the agent's beliefs coherent.

Iterated Revision

The sequences model handles iterated revision in straightforward fashion. At every stage of the representation, we have a new sequence, and we are concerned with the status of inference from that sequence. There is no change in the representation of the beliefs after revision: iterated revision becomes a trivial notion. All the information that is needed to guide revision or inference from that sequence is available in the sequence itself.

Prioritized Inference

With this approach, we want to distinguish between different inconsistent belief sequences: we want it to be the case that the set of inferences that can be drawn from two sequences σ_1, σ_2 could differ even though $Cn[\![\sigma_1]\!] = Cn[\![\sigma_2]\!] = K_\perp$. We use the *maxiconsistent* approach to define *prioritized* inference on a belief sequence σ. This method employs a restricted part of σ: a maxiconsistent subset (σ itself can be an inconsistent set). The elements of a belief sequence $\sigma : \beta_1, \ldots, \beta_n$ are *reorganized* into a new sequence $\delta_1, \ldots, \delta_n$ by the context created in answering the query 'γ?' (γ is a formula whose deducibility from that sequence is to be determined). Answering a query causes a reshuffling of the agent's belief sequence; the temporal ordering on the formulae in the sequence is affected by a relevance ordering. We describe this method for answering queries as follows–consider a formula γ in its smallest language L_γ. We construct a maxiconsistent subset $\Gamma_{\langle\sigma,k,\gamma\rangle}$ (of k-relevant to γ formulas) of σ. The construction of this set is regulated by the ordering $\prec$ that γ creates on σ.

DEFINITION 35 *$\forall \delta_i, \delta_j \in \sigma, \delta_i \prec \delta_j$ if a) $\exists r$, δ_i is r-relevant to γ and δ_j is not r-relevant to γ (i.e., δ_i is more relevant to γ than δ_j) or b) δ_i, δ_j are equally relevant but $j < i$, i.e., δ_i is more recent than δ_j.*

The $\delta_1, \ldots, \delta_n$ are the $\beta_1, \ldots, \beta_n$ under this order. This ordering assigns higher ranks to more relevant formulas. When two equally relevant formulas have to be arranged in the sequence, we assign a higher priority to the more recent formula. Both the relevance of the formula and the temporal order of the receipt of the formula play a part in its position in the ordering. Clearly, if if $\delta_i \prec \delta_j$ and $\delta_j \prec \delta_k$ then $\delta_i \prec \delta_k$. We now describe the construction of maxiconsistent subsets of the belief sequence. In the definition below, Γ^n is the set $\Gamma_{\langle\sigma,k,\gamma\rangle}$ referrred to above (we drop the subscripts to improve readability).

DEFINITION 36 $\Gamma^0 = \emptyset$,
$$\Gamma^{i+1} = \begin{cases} \Gamma^i & \textit{if } \Gamma^i \vdash \neg\delta_{i+1} \textit{ or if } \delta_{i+1} \textit{ is not } k\textit{-relevant to } \gamma \\ \Gamma^i \cup \{\delta_{i+1}\} & \textit{otherwise} \end{cases}$$
$\Gamma_{\langle\sigma,k,\gamma\rangle} = \bigcup_{i=1}^{n} \Gamma^i_{\langle\sigma,k,\gamma\rangle} = \Gamma^n$.

We check formulas for addition to $\Gamma_{\langle\sigma,k,\gamma\rangle}$ in order of their decreasing relevance (increasing k-relevance) to γ. The lower the level of relevance allowed (i.e.. the higher the value of k), the larger the part of σ considered. When making an inference. we consider the most relevant (and recent) of our beliefs before considering other. less relevant (and older) beliefs. At each step of the construction, we maintain the consistency of the set Γ^n. We now define the inference operation $\vdash_k$.

DEFINITION 37 $\sigma \vdash_k \gamma$ *iff* $\Gamma_{\langle\sigma,k,\gamma\rangle} \vdash \gamma$

A formula γ can be inferred (at some level of relevance k) from a belief sequence σ if it is classically entailed by the maxiconsistent subset (of k-relevant to γ formulas) of σ. We consider the most immediately relevant, temporally and contextually, formulas in this inference and none other. Once $\Gamma_{\langle\sigma,k,\gamma\rangle}$ has been constructed, the agent can answer the query γ with a definite response:

- If $\Gamma_{\langle\sigma,k,\gamma\rangle} \vdash \gamma$, then answer 'yes'.
- If $\Gamma_{\langle\sigma,k,\gamma\rangle} \vdash \neg\gamma$ then answer 'no'.

- Otherwise, answer 'no information'.

Even if the sequence σ is inconsistent, the agent is able to give consistent answers to every query.

The notion of inference thus defined has some desirable features. As an example. suppose our belief sequence is initiated by first being told p and then $\neg p \wedge \neg q$. Clearly. we will no longer answer 'yes' to p. However, if we are now told $p \vee q$. this new information overrides $\neg p \wedge \neg q$. Thus, the maxiconsistent set in question. $\Gamma_{<\sigma.0.p>}$ is $\{p \vee q, p\}$ and the query p? will now be answered in the affirmative. This is plausible since the latest information decreases the reliability of $\neg p \wedge \neg q$ and the original information p regains its original standing. These variations are not easily accomodated within traditional, AGM-based frameworks for belief revision.

Prioritization of Directly relevant formulas

Further depth can be introduced into the method provided above by introducing a prioritization amongst formulas based on the *amount* of language overlap between directly relevant formulas. Such a prioritization $\preceq_{|L|}$ might be defined as follows: let β_1, β_2 be formulas directly relevant to γ: $\beta_1 \preceq_{|L|} \beta_2$ if $|L_{\beta_1} \cap L_\gamma| > |L_{\beta_2} \cap L_\gamma|$. Under this prioritization scheme, we further refine the notion of relevance in order to pick the *most* relevant set of formulas to the query. This scheme aims to capture the intuition that formulas that share more symbols are more relevant to each other than those that share fewer symbols. Such a measure however, is open to objections that propositions that share fewer symbols might be more relevant than those that share more symbols *depending upon the symbols shared.* As an example, consider our intuition that if two formulas mention *aardvarks*, they are (most likely) more relevant to each other than if they mentioned *cats*. This notion is further refined then, by the following heuristic: if symbols shared by formulas occur frequently in the language. then there is a lesser likelihood that the formulas are more relevant to each other than if they share symbols that occur with less frequency in the language. This provides for a notion of *degree of relevance* amongst directly (and only for directly) relevant formulas.

DEFINITION 38 *Consider formulas* α, β, γ *such that* $L_\alpha \cap L_\beta = \Delta_1, L_\alpha \cap L_\beta = \Delta_2$ *i.e.,* α, β *and* α, γ *are directly relevant. For a belief sequence* σ, *let*

$$|\{\psi | \psi \in \sigma \wedge \Delta_1 \subseteq L_\psi\}| = m$$

$$|\{\phi | \phi \in \sigma \wedge \Delta_2 \subseteq L_\phi\}| = n$$

If then, $m > n$, *then the degree of direct relevance of* α, β *w.r.t* σ $drel(\alpha, \beta, \sigma)$ *is less than the degree of direct relevance of* α, γ *w.r.t* σ *i.e.,* $drel(\alpha, \beta, \sigma) < drel(\alpha, \gamma, \sigma)$ *where* $drel(\alpha, \beta, \sigma) =$

$$\frac{|\sigma|}{\{\psi | \psi \in \sigma \wedge \Delta_1 \subseteq L_\psi\}}$$

In keeping with the ternary notion of relevance developed in this study, the degree of direct relevance between two formulas is dependent on a sequence σ and is not just a matter of syntactic or logical relations between the formulas in question. The effect of the definition above on the ordering created in the sequence by the query α will affect the definition of the inference operator $\vdash_k$. We defer such a development to a more leisurely discussion.

Answer sets and Consequence relations

We define a set of consequences of a sequence σ, $C_k(\sigma)$, at a particular level of relevance, k:

DEFINITION 39 $C_k(\sigma) = \{\gamma | \sigma \vdash_k \gamma\}$

The following defines the set of consequences of a sequence σ at all levels k:

DEFINITION 40 $C(\sigma) = \{\gamma | \exists k, \sigma \vdash_k \gamma\} = \bigcup C_k(\sigma)$

Thus we could have $Cn[\![\sigma_1]\!] = Cn[\![\sigma_2]\!] = K_\perp$ but $C(\sigma_1) \neq C(\sigma_2)$. An answer set as follows will always be consistent for any fixed set of subject matters:

OBSERVATION 12 *If Σ is a set of propositional atoms, then* $\{\alpha | L_\alpha = \Sigma\} \cap C_k(\sigma) \not\vdash \perp$.

That is, the agent's responses to a particular subject matter are guaranteed to be consistent by the query answering scheme.

Properties of $\vdash_k$

PROPOSITION 3 *The inference procedure defined above is monotonic in k, the level of relevance i.e., if $\sigma \vdash_k \gamma$ then $\sigma \vdash_{k+1} \gamma$.*

There is no loss of information (with respect to a particular query) in stopping at a particular value of k-relevance since any formula derivable at that point will be derivable later as well. This property has the value of being sensitive to the resources available to us during a particular inference operation since we can choose to stop as soon as we get an answer. Ergo, once the agent has inferred from its beliefs at a particular level of relevance that a certain belief follows, our construction makes it the case that the same belief will also be believed at a lower level of relevance (i.e., higher values of k-relevance).

PROPOSITION 4 *The inference procedure is non-monotonic in expansions of a belief sequence i.e., if $\sigma \sqsubseteq \sigma'$ and $\sigma \vdash_k \gamma$ then it is not necessarily the case that $\sigma' \vdash_k \gamma$.*

The above is obvious by the definition of the inference procedure and renders this procedure a plausible representation of a belief revision procedure. As a reminder, a non-monotonic inference operation defined on an explicit representation of beliefs can be said to serve as a belief revision operation since it corresponds to an agent dropping inferences that might have been made before the receipt of new information. We now turn to other properties, often held to be plausible properties for inference relations and belief revision operators:

PROPOSITION 5 *The following properties hold for the process of revision defined on sequences:*

- *Weak Inclusion: If* $\gamma \neq \perp$ *then* $\sigma * \gamma \vdash_k \gamma$

The following hold under the condition that $L_\alpha = L_\beta$*:*

- *Weak (or Cautious) Monotonicity:*

$$\frac{\sigma \vdash_k \alpha, \sigma \vdash_k \beta}{\sigma * \alpha \vdash_k \beta}$$

- *Rational Monotonicity:*

$$\frac{\sigma \not\vdash_k \neg\alpha, \sigma \vdash_k \beta}{\sigma * \alpha \vdash_k \beta}$$

- *Weak Cut:*

$$\frac{\sigma * \alpha \vdash_k \beta, \sigma \vdash_k \alpha}{\sigma \vdash_k \beta}$$

- *Adjunction:*

$$\frac{\sigma \vdash_k \alpha, \sigma \vdash_k \beta}{\sigma \vdash_k \alpha \wedge \beta}$$

- *Right Weakening:*

$$\frac{\sigma \vdash_k \alpha, \alpha \vdash \beta}{\sigma \vdash_k \beta}$$

That Weak Monotonicity fails in general (i.e., unless $L_\alpha = L_\beta$) is easily demonstrated by the following example: let $\sigma = [p, \neg p \wedge \neg q, q]$. Now, $\forall k, \sigma \vdash_k p \wedge q$, but also, $\sigma \vdash_k \neg p$. However, $\sigma * (p \wedge q) \not\vdash_k \neg p$. Similar examples can be constructed for the other rules mentioned above. If two queries α, β are asked on the same occasion, the query answering procedure would use $L = L_\alpha \cup L_\beta$ as the language for determining direct relevance. The answers thus generated will be compatible with each other. While some of the properties indicated above might seem unreasonable, they become necessary since we are modeling real agents who often have significant logical deficiencies.

'Contraction'

In general, 'contraction' in the case of sequences works on the principle that we have to have a *reason* for contraction. New formulas block the derivation of older formulas and in doing so, provide a means for contraction. What does this explicit notion of contraction tell us about the notion of contraction as used in standard AGM models? It informs us that contraction needs to be done on a principled basis. We do not just contract our beliefs; we contract for a reason, so that to contract by α is to *revise* by some information that implies $\neg\alpha$. Such a notion meshes well with our intuitions about the notion of contraction. In our model it is possible to lose α without acquiring $\neg\alpha$, that is, the model accommodates belief change scenarios in which belief in a proposition is given up without corresponding belief in the negation of the proposition being present i.e., it accommodates a move from acceptance of a proposition to holding that same proposition in suspense. For example, consider the sequences $\sigma = p \wedge q$ and $\sigma * (\neg p \vee \neg q)$. The revised sequence does not answer 'yes' to p but neither it does answer 'yes' to $\neg p$. In the pilot example presented in Chapter 1, the pilot drops his older belief that there is a hostile craft in front of him only when he sights the markings of the friendly air force (this new information overrides the older and is directly relevant to his beliefs about aircraft sightings), and he drops his beliefs that there are enemy aircraft in the region upon receipt of newer information that all enemy planes have left.

There is no clear notion of minimal change built into the sequences model since there is no precise counterpart of the contraction operation in the sequences model (there is no precise counterpart of the expansion operation either). However the reason based contraction of the sequences model provides one way of capturing the notion of minimal change. In this regard, there are strong similarities between the sequences model and base revision schemes. Furthermore, since only relevant formulas from a belief sequence can contribute to a particular query's derivation being blocked, the sequences model makes no room for any kind of trivial update either.

Equivalence of Belief Sequences

Given the notion of inference defined above, we say that belief sequences which yield the same answers to all queries are *equivalent*. Consider the following example: $\sigma_1 = [\neg p \wedge \neg q], \sigma_2 = [p, \neg p \wedge \neg q]$. Neither σ_1 or σ_2 implies p as a conclusion. Revising by the formula $p \vee q$, however, restores p as a conclusion for σ_2 while σ_1 gives 'no information' as an answer. So, while these sequences were initially equivalent, revision by the same formula changed their answer set. Given this observation, we would like to define a strong notion of equivalence which is unaffected by revision.

DEFINITION 41

- *Sequences σ_1, σ_2 are k-equivalent iff $\forall \gamma,\ \sigma_1 \vdash_k \gamma \Rightarrow \sigma_2 \vdash_k \gamma$.*
- *Sequences σ_1, σ_2 are strongly k-equivalent iff $C_k(\sigma_1 * \alpha_1 * \ldots \alpha_n) = C_k(\sigma_1 * \alpha_1 * \ldots \alpha_n)$ for all sequences of revisions $\alpha_1, \ldots, \alpha_n$.*

- *σ_1, σ_2 are weakly k-equivalent iff:*

  $\sigma_1 \vdash_k \gamma \Rightarrow \sigma_2 \vdash_k \gamma$ *or if the query answering scheme provides 'no information' from σ_2 i.e., a neutral answer.*

  $\sigma_1 \not\vdash_k \gamma \Rightarrow \sigma_2 \not\vdash_k \gamma$ *or if the query answering scheme provides 'no information' from σ_2.*

We conjecture that the notion of strong equivalence is testable via a single revision. that is, if σ_1 and σ_2 are not strongly k-equivalent, then there is a γ such that $\sigma_1 * \gamma$ and $\sigma_2 * \gamma$ are not weakly k-equivalent. In general, we can think of revision of sequences as an equivalence disturbing operation in that successive revisions can reduce equivalent sequences to weakly equivalent sequences and then to inequivalent sequences. An obvious question is, how can all sequences be trimmed or reduced to their shortest equivalent form? The task of reducing sequences to their simplest equivalent form is most likely computationally non-trivial.

Sequence 'Trimming'

A time and space saving strategy is for the agent to be able to drop formulas that are not likely to play an active role in any inference. The condition for a sequence σ to be 'trimmed' by dropping formulas is in terms of language overlap and the logical relationships between formulas in the sequence.

OBSERVATION 13 *For a sequence σ, if α precedes β, in σ, $L_\alpha \subseteq L_\beta$ and $\beta \vdash \neg\alpha$ then we can replace σ by $\sigma' = \sigma - \alpha$. Similarly if $\alpha \Leftrightarrow \beta$.*

This allows for reduction in the size of the sequence σ and reduces the complexity of the inference procedure. The equivalence of σ and σ' is *not a strong equivalence*, however: σ, σ' are weakly k-equivalent.

Comparison with B-structures Query Answering

In the B-structures model, we presented a method for answering queries that allowed the answer $\top$, i.e., (*over-defined* or *inconsistent*). In the current model we block the possible inconsistency of Γ^n, thus preventing an inconsistent answer. The query answering procedures of the B-structure model will agree with our method when the former gave 'yes' or 'no' answers. As an option to reduce the complexity of the procedure for answering queries, we could relax the consistency condition on the construction of $\Gamma_{\langle\sigma,k,\gamma\rangle}$ and avail ourselves of a four-valued logic as in the case of the B-structures model. The B-structure query answering method corresponds to $\vdash_0$ i.e.. for Γ^0 constrained to consider only directly relevant formulas.

Conformance with AGM Axioms

It is natural to ask whether our inference conforms to AGM-like postulates. We cannot apply the results in [38] that show that maxiconsistent inference on a belief sequence is rational; our inference procedure takes into account relevance relations that cause a reordering of our sequence depending on the formula whose inference we are testing. However, there are some AGM-like properties that do hold. We present axioms roughly analogous to the AGM axioms.

(σ^* 1) $\sigma * \gamma$ is a belief sequence.

(σ^* 2) $\gamma \in C_k(\sigma * \gamma)$

(σ^* 3) If $\vdash \alpha \leftrightarrow \beta$ then $\forall\gamma$, $\sigma * \alpha \vdash_k \gamma$ iff $\sigma * \beta \vdash_k \gamma$.

Since α, β are expressed in their smallest language, this result is immediate.

Conformance with Lehmann Axioms

A comparison with the Lehmann axioms shows that our method of inference defined above does not conform to them. The only one that the method conforms to is **I2**. The reason for this is that in addition to the temporal ordering of the belief set.

we have also imposed a relevance ordering. The combination of these two orderings ensures that formulas whose inference was possible at one stage in the history of the agent could be blocked at later stages in the history of the agent by newer information.

Complexity of the Inference Procedure

In the methods defined above, there are two sources of complexity. The first one involves the calculation of the smallest language which is a co-**NP** complete problem ([55]), and the second one involves the complexity of checking the consistency of the set Γ_i at each step of the construction of the set Γ^n. The answering method is dependent upon first calculating the relevance relation exhaustively for the entire sequence. So, the entire sequence must be checked for the possible existence of a directly relevant formula and then the trimmed sequence is checked again for 1-relevant formulas and so on. At each step, the sequence shrinks. Since each stage of the construction involves a consistency check, the complexity of the procedure is polynomial with an **NP** oracle. The reduction of this complexity requires further investigation.

Conclusion

The sequences model just described possesses the advantage of being easy to implement as we only need to keep track of the sequence. We do not drop any incoming information; it might be useful in later stages of revision. How psychologically and philosophically plausible is such a model? Perhaps its most distinctive feature is that in the case of sequences of formulas, the order of the incoming information plays a role in the inference procedures. The method above pays adequate heed to the psychological plausibility of this fact. The inference procedure gains further plausibility by being relevance sensitive: resource bounded agents are unable to keep track of all available information and must rely on notions like relevance and context to aid in inference.

The sequences model is explicitly inconsistency tolerant via the notion of inference that blocks the derivation of arbitrary beliefs from an inconsistent set of beliefs. While the current definition of the query answering procedure for the sequences model works only with a classical truth value space for its responses, it would be easy to build into the model a four valued space; we could relax the maxiconsistency condition in the inference procedure. The set Γ_n would then be constructed by simply collecting all formulas up to the level of relevance specified and then testing for the derivation the query in question. This would have the added advantage of improving the tractability of the procedure with the tradeoff being that the agent would, under certain levels of relevance, not provide definite answers. Since epistemic states in the sequences model are the non-monotonic consequences of the inference operation that we have provided, there would be cases in which a coherent epistemic state could not be captured. This would be no worse however, than the situation in the case of the B-structures model. In that model, as in the sequences model, the understanding is that the construction of the inference operation is so as to model coherent epistemic states *most of the time*. Working with the maxiconsistent construction has the advantage that it leaves the door open for a possible notion of models for a belief sequence and therefore, a notion of semantic consequence. Such a notion, however, requires further investigation and is deferred for the time being[2].

[2]Conversely, in the query answering procedure for B-structures, a maxiconsistency requirement could be built in i.e., the set Γ_α could be constructed using directly relevant formulas to the query from the B-structure with the constraint that formulas inconsistent with Γ_α would be rejected. This would facilitate a design of a notion of semantic consequence for inference from B-structures as well. As with the sequences model, we defer such an investigation for the time being. It is unclear too, whether such a notion would add to our understanding of the epistemic states represented by the

The definition of relevance used in the sequences model (and the B-structures model) can be thought of as a simplification of the notion of cognitive relevance. A richer and more psychologically plausible notion of relevance will require richer (higher-order) languages and an active investigation of the role of contexts being triggered by conversations. Another possibility (as suggested by Fitting) is to work with a purely stipulative notion of relevance i.e.. specify at the very outset which propositions are relevant to others in the belief representation and then work with that classification.

models under study.

Appendix A

Appendix: AGM Constructions

A.1 Constructions for AGM revision

Since the AGM axiomatizations do not tell us *how* to revise a belief set. we need constructive mechanisms for doing so i.e., a method that tells what our belief set *is* after revision and not just what its form is. I now briefly describe some explicit constructions devised within the AGM model beginning with those suggested by AGM themselves. Work in belief revision on the minimal change model has concentrated on minimizing the number of beliefs given up during the *contraction* operation: if one takes the Levi identity as primary this approach follows. To this end, operations using so-called *selection functions* were defined in [2]. I present a brief overview of some of the techniques presented in this seminal paper. We begin with the concept of a maximal subset of the belief set that fails to imply a belief α; the intuition being that this is what we would like our contracted belief set to look like after dropping α.

DEFINITION 42 [35] *A belief set K' is a maximal subset of K that fails to imply α if and only if:*

1. $K' \subseteq K$

2. $\alpha \notin K'$

3. For any $\beta \in \mathcal{L}$, if $\beta \in K$ and $\beta \notin K'$, then $\beta \rightarrow \alpha \in K'$

The set of all maximal subsets of K failing to imply α is denoted $K \perp \alpha$. Given the existence of this set, we apply a selection function γ that returns its 'best' element; γ is known as a *maxichoice selection function*. The contraction of K by α is then defined as follows:

DEFINITION 43 [2] $K_\alpha^{\dot{-}} = \begin{cases} \gamma(K \perp \alpha) & \textit{if } K \perp \alpha \neq \emptyset \\ K & \textit{otherwise} \end{cases}$

While such a definition satisfies the AGM postulates it has the undesirable property of retaining *too much* information: the agent has opinions on the truth or falsity of every proposition in the language. A similar definition for a *full meet selection function* that returns only those elements that occur in all elements of $K \perp \alpha$ goes in the other extreme retaining *too little* information (while satisfying the AGM postulates):

DEFINITION 44 [2] $K_\alpha^{\dot{-}} = \begin{cases} \bigcap(K \perp \alpha) & \textit{if } K \perp \alpha \neq \emptyset \\ K & \textit{otherwise} \end{cases}$

Such a function is at clear odds with the Principle of Minimal Change. To overcome these problems, a *partial meet selection function* returns the *set of 'best' subsets* of $K \perp \alpha$ and is used to define a *partial meet contraction function* as follows:

DEFINITION 45 [2] $K_\alpha^{\dot{-}} = \begin{cases} \bigcap \gamma(K \perp \alpha) & \textit{if } K \perp \alpha \neq \emptyset \\ K & \textit{otherwise} \end{cases}$

A representation theorem then connects the definition above to the AGM postulates:

THEOREM A.1 [2] *Let K be a belief set and $\dot{-}$ be a contraction function. Then $\dot{-}$ is a partial meet contraction function over K if and only if it satisfies postulates $(K^{\dot{-}}1) - (K^{\dot{-}}6)$ for contraction over K.*

An ordering is imposed over the elements of $K \perp \alpha$ to pick out the most preferred elements; the ordering is required to be transitive and connected.

Model Based Revision

Winslett [109] , Katsuno and Mendelzon [59] and Dalal [14] suggest a semantical model of belief revision. In this approach, revisions of an agent's epistemic state are to be performed at the *model-theoretic* level rather than at the level of sentences associated with the epistemic state. A belief set is viewed as a set of models that satisfy a particular set of explicit beliefs and belief revision is done in a way that selects the *models* that satisfy the resultant set of sentences. The new set of models is to differ *minimally* from the models of the original belief base. To make the notion of minimal difference coherent, we must say something about how models are to be compared. For this, a notion of *ordering amongst models* is developed as described below.

An *interpretation* of L is a function from the set of propositional atoms P to the set of truth values: $I : P \times \{T, F\} \rightarrow \{T, F\}$. The set of all interpretations is $\mathcal{I}$; the models of a sentence α is the set $\{I | I \in \mathcal{I}, I(\alpha) = T\}$ with I extended to all formulas. Now, using an ordering over the set of all interpretations, we determine which interpretations should be interpreted as models of a theory $K * \alpha$ and therefore, determine $K * \alpha$. The intuition here is that some models of α but not of K are closer to models of K than others. In order to conform to the axioms $(K^*3), (K^*4)$, the models of K are taken to be the 'smallest' elements in I. Katsuno and Mendelzon define a *persistent* ordering $\leq_K$ over the set of interpretations I of L for each belief set K as follows:

DEFINITION 46 [58] *1. If $I \in Mod(K)$, then $I \leq_K J$, for all interpretations J.*
2. If $I \in Mod(K)$, and $J \notin Mod(K)$, then $I <_K J$.

If Int is a set of intepretations of L, $Min(Int, \leq_K)$ denotes the set of interpretations in I which are minimal in Int with respect to $\leq_K$. $K^*\alpha$ then, is the set which has exactly $Min(Mod(A), \leq_K)$ as its set of models. The following theorem connects the AGM postulates and the definition provided above:

THEOREM A.2 [58] *A revision function* $*$ *satisfies* $(K^*1) - (K^*6)$ *if and only if there exists a persistent total pre-ordering* $\leq_K$, *such that* $Mod(K^*\alpha) = Min(Mod(\alpha), \leq_K)$.

As an example of the model based approach, Dalal [14] measures the distance between two finite interpretations, $I, J, dist(I, J)$ in terms of the number of propositional variables that have different truth values in I and J. We can then construct a *persistent* ordering of interpretations as follows:

DEFINITION 47 [14] $dist(Mod(K), I) = min\{dist(J, I) : J \in Mod(K)\}$. *Then the ordering* $\leq_K$ *is defined as* $I \leq_K J$ *if and only if* $dist(Mod(K), I) \leq dist(Mod(K), J)$

The $\leq_K$ defined above can be shown to be a persistent ordering. By the theorem above it follows that the revision function defined above satisfies the AGM postulates. I now present in some detail, the most well known of the representation results that connect model based revision schemes with the AGM axioms.

Grove's System of Spheres

Grove [44] considers a *system of spheres model* that is largely inspired by David Lewis' *spheres semantics* [67] for counterfactual reasoning. Grove takes maximally consistent sets of formulae (maximal extensions of consistent complete theories) to be 'possible worlds' and places an ordering $M_{\mathcal{L}}$ over the set of all possible worlds. The possible worlds consistent with any set K, (denoted by $[K]$) are defined as follows:

DEFINITION 48 [44] $[K] = \begin{cases} m \in \mathcal{M}_{\mathcal{L}} : K \subseteq m & \text{if } K \neq K_{\perp} \\ \emptyset & \text{otherwise}^1 \end{cases}$

The possible worlds consistent with an arbitrary formula α are denoted $[\alpha]$ which is defined as $[\alpha] = \{m \in \mathcal{M}_{\mathcal{L}} : \alpha \in m\}$. A function $th : 2^{\mathcal{M}_{\mathcal{L}}} \rightarrow \mathcal{K}$ (which corresponds to a theory) is then defined. For any $X \subseteq \mathcal{M}_{\mathcal{L}}$, we have:

DEFINITION 49 $th(X) = \begin{cases} \bigcap\{m \in X\} & \text{for } X \subseteq \mathcal{M}_{\mathcal{L}} \text{ and } X \neq \emptyset \\ K_{\perp} & \text{if } X = \emptyset \end{cases}$

The function th as expected, has the following properties:

LEMMA 3 *$th([K]) = K$ for all belief sets K if the underlying logic is compact.*
$th(X) \neq K_{\perp}$ if and only if X is nonempty.
For any sentence $\alpha \in \mathcal{L}$ and $X \subseteq M_{\mathcal{L}}, th(X \cap [\alpha]) = Cn(th(X) \cup \{\alpha\})$.
For $X, X' \subseteq M_{\mathcal{L}}$, if $X \subseteq X'$, then $th(X' \subseteq (X)$.
For $K, K' \in K$. if $K \subseteq K'$, then $[K']$.

A sphere is defined to be a set of possible worlds. A system of spheres centered on K is defined as a complete quasi-ordering over sets of sets of possible worlds where $[K]$ is the innermost sphere and $M_{\mathcal{L}}$ is the outermost:

DEFINITION 50 [44] *Let S be any collection of subsets of $M_{\mathcal{L}}$ (i.e., sets of sets of possible worlds). We call $\mathcal{S}$ a system of spheres, centered on X for some subset $X \subseteq M_{\mathcal{L}}$, if it satisfies the following conditions:*
(S1) $\mathcal{S}$ is totally ordered by $\subseteq$; that is, if $U, V \in S$, then $U \subseteq V$ or $V \subseteq U$.
(S2) X is the $\subseteq$-minimum of $\mathcal{S}$ (i.e., $X \in \mathcal{S}$ and if $U \in \mathcal{S}$, then $X \subseteq U$.
(S3) $M_{\mathcal{L}}$ is in $\mathcal{S}$ (the maximal element of $\mathcal{S}$).
(S4) If $\alpha \in \mathcal{L}$, and there is any sphere in $\mathcal{S}$ intersecting $[\alpha]$, then there is a smallest sphere in $\mathcal{S}$ intersecting $[\alpha]$ (there is a sphere $U \in \mathcal{S}$ such that $U \cap [\alpha] \neq \emptyset$ and $V \cap [\alpha] \neq \emptyset$ implies $U \subseteq V$ for all $V \in \mathcal{S}$.

The crucial part of the definitions above is ($S4$) since it assures that if any formula α has worlds intersecting $M_{\mathcal{L}}$, then there is a smallest sphere (or innermost) that intersects $[\alpha]$. Let $C_{\mathcal{S}}(\alpha)$ be such a sphere. If $[\alpha]$ does not intersect any sphere in $\mathcal{S}$, then $C_{\mathcal{S}}(\alpha) = M_{\mathcal{L}}$. Now, we can associate, with any system of spheres $\mathcal{S}$ centered on $[K]$, a function $f_s : \mathcal{L} \rightarrow 2^{M_{\mathcal{L}}}$ defined as follows:

For any $\alpha \in \mathcal{L}$, $f_s(\alpha) = [\alpha] \cap c_s(\alpha)$

It is clear that the function f_s is selecting those α-worlds in $M_{\mathcal{L}}$ that are closest to $[K]$. The revision operator is now defined in straightforward fashion:

DEFINITION 51 $[K^*_\alpha] = f_S(\alpha)$

The operator above ensures that the worlds corresponding to a revision of K by α are exactly those α-worlds (a world in which α holds; $[\alpha]$ is the set of all α-worlds) closest to $[K]$. This is in conformance with the spirit of the requirement of Minimal Change where minimality is defined to be 'closeness' to the original belief set $[K]$. The spheres method picks out a set of worlds that is as close as possible to the original epistemic state; the minimum deviation from the original is achieved in this fashion. Grove then obtains the following representation theorems:

THEOREM A.3 *Let $\mathcal{S}$ be any system of spheres in $M_{\mathcal{L}}$ centered on $[K]$ for some belief set K in $\mathcal{K}$. If one defines, for any $\alpha \in \mathcal{L}$, K^*_α to be $th(f_s(\alpha))$, then postulates $(K*1)$-$(K*8)$ are satisfied.*

THEOREM A.4 *Let $* : \mathcal{K} \times \mathcal{L} \rightarrow \mathcal{K}$ be any function satisfying postulates $(K*1)$-$(K*8)$. Then for any (fixed) belief set K, there is a system of spheres on $M_{\mathcal{L}}$, $\mathcal{L}$, $\mathcal{S}$ say, centered on $[K]$ and satisfying $K^*_\alpha = th(f_s(\alpha))$ for all $\alpha \in \mathcal{L}$.*

The belief revision operations can now be provided a straightforward semantics. In the case of expansion, when $\neg\alpha \notin K$, we have $[K] \cap [\alpha] \neq \emptyset$. This means that the closest α-worlds are in the innermost sphere $[K]$ and the worlds consistent with the expanded epistemic state are given by $[K^{\dot{+}}_{\alpha}] = [K] \cap [\alpha]$. When $\neg\alpha \in K$, we have $K^{\dot{+}}_{\alpha} = K_{\perp}$. In this case $[K] \cap [\alpha] = \emptyset$ and so we get $[K^{\dot{+}}_{\alpha}] = [K] \cap [\alpha]$. In similar fashion (and by using the Harper identity) we obtain an expression for the contraction operation as follows: $[K^{\dot{-}}_{\alpha}] = [K] \cup f_s(\neg\alpha)$

Epistemic Entrenchment

In the AGM model, since an agent's epistemic state is given by the belief set, information about strength of beliefs cannot be expressed. This means that the same revision can result in different belief sets if we consider alternative strengths for the beliefs in the belief set. For example, take our belief set, $K = Cn(p \wedge q)$, and the new information $\alpha = \neg p \vee \neg q$. Now, if our belief in p is stronger than our belief in q then $K^{*}_{\alpha} = \{p\}$. If however, our belief in q is stronger than belief in p then K^{*}_{α} $= \{q\}$. If we consider the information about strength of beliefs to be included in the operator itself, then the assumption that the same operator $*$ persists in iterated revisions is mistaken: there is no reason to believe that the relative strength of beliefs is maintained after we learn a formula. The assumption then, that a revision operator captures the agent's general epistemic strategy is mistaken: we need to conceive of revisions as being carried out a by a new operator in each case. We need to supplement belief sets with the relative strength of beliefs[2]: what we mean by an epistemic state is not captured by the notion of a belief set. A motivation for the choice of

[2]This invites talk of probabilities: any attempt to formalize the notion of strength of beliefs is probably committed to some such notion.

beliefs to be dropped is given by the notion of *epistemic entrenchment* introduced by Gärdenfors and Makinson[36]. The guiding spirit behind entrenchment is that *the more entrenched the belief, the less likely it is to be dropped during any revision operation*: this formalizes the notion of the strength of beliefs[3]. Entrenchment is clearly a cluster concept: formulas differ in entrenchment depending upon their relative informational value, their justificatory pedigree and so on. This information is not present in a belief set by itself: it must be provided as an additional input to the revision process. This enriches an epistemic state considerably; rather than just the sparseness of a belief set, we now have information on the relative strength of beliefs. Formally, an entrenchment relation can be defined as follows (it will be noticed that an epistemic entrenchment is a total preorder over formulas such that non-beliefs are minimally entrenched and tautologies are maximally entrenched):

DEFINITION 52 [36] *An entrenchment ordering $\leq$ over L satisfies the following conditions*

- *(EE1) For any $\alpha, \beta, \gamma \in \mathcal{L}$, if $\alpha \leq \beta$ and $\beta \leq \gamma$ then $\alpha \leq \gamma$ (Transitivity)*
- *(EE2) For any $\alpha, \beta \in \mathcal{L}$, if $\{\alpha\} \vdash \beta$ then $\alpha \leq \beta$ (Dominance)*
- *(EE3) For any $\alpha, \beta \in \mathcal{L}$, either $\alpha \leq \alpha \wedge \beta$ or $\beta \leq \alpha \wedge \beta$ (Conjunctiveness)*
- *(EE4) When $K \neq K_{\perp}, \alpha \notin K$ iff $\alpha \leq \beta$ for all $\beta \in \mathcal{L}$ (Minimality)*
- *(EE5) If $\beta \leq \alpha$ for all $\beta \in \mathcal{L}$, then $\vdash \alpha$ (Maximality)*

[3] I thank Henry Kyburg for pointing out that probabilistic talk is almost inevitable once talk of strength of beliefs is introduced. For the present purposes, and since I introduce entrenchment here merely to survey a kind of construction possible for AGM postulates, I steer clear of any such probabilistic considerations.

The first condition simply states that entrenchment is transitive. The dominance condition says that for any two formulae in the language, the stronger formula is more entrenched: if one of the two formulae must be given up, the less entrenched formula will be given up first. The conjunctiveness condition says that for any two formulae in the language, either formula is more entrenched than the conjunction of the two: giving up $\alpha \wedge \beta$ can be accomplished by giving up either α or β. The minimality condition says that non-beliefs are minimal elements in this ordering while the maximality condition says that logical truths are the hardest to give up. This last condition might be felt to be implausible in some circumstances i.e., I might be more willing to give up belief in some logical truths than to give up the belief that my parents are not my biological parents and that I was adopted. If this is just understood as some sort of pathological stubborness on my part, then all is well.

Any ordering satisfying the first three properties above is called an *expectations ordering.* Some properties of expectation orderings[4] are as follows[5]:

- $\alpha \leq \beta$ or $\beta \leq \alpha$.
- If $\beta \wedge \gamma \leq \alpha$, then $\beta \leq \alpha$ or $\gamma \leq \alpha$.
- $\alpha < \beta$ iff $\beta < \beta$.
- If $\gamma \leq \alpha$ and $\gamma \leq \beta$, then $\gamma \leq \alpha \wedge \beta$.
- If $\alpha \leq \beta$, then $\alpha \leq \alpha \wedge \beta$.

An epistemic entrenchment ordering $\leq$ for a particular belief set K may be constructed from a contraction function $\dot{-}$ as follows:

[4]Foo provides additional properties of epistemic entrenchments and expectation orderings in [28].
[5]Gärdenfors and Makinson, [36].

$(C \leq) : \alpha \leq_{\dot{-}} \beta$ iff $\alpha \notin K^{\dot{-}}_{\alpha \wedge \beta}$ or $\vdash \alpha \wedge \beta$.

The condition states that β is at least as epistemically entrenched as α whenever α is removed from K in contracting K by $\alpha \wedge \beta$. To contract by $\alpha \wedge \beta$, we need to remove one of α or β; given that α has been removed, we can infer that it cannot be strictly more entrenched than β. When α and β are both tautological, they are equally entrenched. Conversely, it is possible to construct a contraction function from an epistemic entrenchment ordering as follows:

$(C \dot{-}) : \beta \in K^{\dot{-}\leq}_{\alpha}$ iff $\beta \in K$ and either $\alpha < \alpha \vee \beta$ or $\vdash \alpha$.

This definition simply states that for β to be in the contracted state, it must have been present in the initial K and either the input α is a logical truth (in which case no change is made to K) or the disjunction of the input with β is more entrenched than the input. Given these definitions, two important representation theorems can be proven:

THEOREM A.5 [36] *Let $K \in \mathcal{K}$ be a belief set and $\leq$ be an epistemic entrenchment over K. If for any $\alpha \in \mathcal{L}$, we define $K^{\dot{-}}_{\alpha}$ using $(C \dot{-})$, then $K^{\dot{-}}1$-$K^{\dot{-}}8$ are satisfied as is the condition $(C \leq)$.*

THEOREM A.6 [36] *Let $\dot{-}: \mathcal{K} \times \mathcal{L} \rightarrow \mathcal{K}$ be any function satisfying $K^{\dot{-}}1$-$K^{\dot{-}}8$. Then for any belief set $K \in \mathcal{K}$, if we define $\leq$ using condition $(C \leq)$, then $\leq$ is an epistemic entrenchment ordering i.e., it satisfies EE1-EE5 and also satisfies condition $(C\dot{-})$.*

A.2 Constructions on Bases

The constructions of partial meet contraction have been extended for bases. Interestingly, the special cases of partial meet contraction (maxichoice contraction and full meet contraction) which give trivial results for belief sets give non-trivial results for bases. In general, the recovery postulate is violated by these constructions as are the supplementary postulates.

I describe partial meet contraction as defined by Nebel since it introduces the notion of a *prioritized base*[6] using *epistemic relevance*. Epistemic relevance is defined in much the same fashion as epistemic entrenchment, as a total preorder with maximal elements on the elements of a belief base:

DEFINITION 53 *An epistemic relevance ordering ($\phi \preceq \psi$) on the elements of a belief base is a reflexive and transitive relation such that if $\phi, \psi \in H$, we have $\phi \preceq \psi$ or $\psi \preceq \phi$. There exists at least one maximal element ϕ. The relation of epistemic relevance introduces an equivalence relation written $\phi \simeq \psi$ as follows:*

$$\phi \simeq \psi \text{ iff } (\phi \preceq \psi \text{ and } \psi \preceq \phi)$$

Intuitively $\phi \preceq \psi$ means that ψ is more 'relevant' or has higher priority than ϕ. A belief base with an epistemic relevance ordering is called a prioritized base (a pair $(H, \preceq)$ which is then used by Hansson and Nebel to construct base revision operations by constructing maximal subsets (of the base) that do not imply a belief to be dropped. A prioritized belief base can be represented by a sequence $(H_1, H_2, H_3, \ldots H_n)$ where the H_i's are equivalence classes under $\simeq$ as defined above.

Let $< H, \preceq >$ be a prioritized belief base then $H' \subseteq H$ is a preferred element of $H \perp \alpha$ if for every $H'' \subseteq H$ and every i such that $H' \cap (H_i \cup H_{i+1} \ldots \cup H_n) \subset H'' \cap (H_i \cup H_{i+1} \cup \ldots H_n)$ it holds that $H'' \vdash \alpha$. Therefore, H' maximizes the set of sentences withdrawn from H constrained by the fact that α is not implied. If

[6]This notion of a prioritized belief base will be useful in understanding an example of the direct revision approach due to Brewka [8].

this sort of contraction is to be performed on a theory K with the base H, we use $\bigcap\{Cn(H' : H'$ is a preferred element of $H \perp \alpha)\}$.

Bibliography

[1] Jonathan Adler. Assertion and three paradoxes of belief. In preparation. 1999.

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

[3] A.R Anderson and Jr. N.D Belnap. *Entailment: The Logic of Relevance and Necessity, Vol. 1.* Princeton University Press, 1975.

[4] Horacio Arlo-Costa and Rohit Parikh. Two place probabilities, beliefs and belief revision: on the foundations of iterative belief kinematics. In *Proceedings of the The Twelfth Amsterdam Colloquium*, 1999.

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

[6] Craig Boutilier. Revision sequences and nested conditionals. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*. pages 519–525. 1993.

[7] Craig Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68:33– 85, 1994.

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

[9] Christopher Cherniak. *Minimal Rationality.* MIT Press, Cambridge. 1986.

[10] Samir Chopra. Konstantinos Georgatos, and Rohit Parikh. Nonmontonic inference on belief sequences (preliminary report). In *Proceedings of the 2nd Pan-Hellenic Symposium on Logic*. Phokion Kolaitis and George Koletsos. eds.. 1999.

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

[12] M.J. Cresswell. *Logics and Languages*. Methuen, London, 1973.

[13] N.C.A. da Costa. On a theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, XVIII(4):621–630, 1974.

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

[15] A.J Dale. The illogic of inconsistency. *Philosophical Studies*. 46:417–425. 1984.

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

[17] Jon Doyle. A truth maintenance system. *Artificial Intelligence*, 12:231–272, 1979.

[18] Jon Doyle. Knowledge, representation and rational self-government. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*. Morgan Kaufmann, 1998.

[19] J. Michael Dunn. Intuitive semantics for first-degree entailments and 'coupled trees'. *Philosophical Studies*, 29:149–168, 1976.

[20] T. Eiter and G. Gottlob. On the complexity of propositional knowledge base revision. *Artificial Intelligence*, 57:227–270, 1992.

[21] Brian Ellis. *Rational Belief Systems*. Rowan and Littlefield. Totowa, 1979.

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

[23] Ronald Fagin, Gabriel M. Kuper, Jeffrey D. Ullman, and Moshe Y. Vardi. Updating logical databases. *Advances in Computing Research*, 3:1–18, 1986.

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

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

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

[27] Jerry Fodor. Modules, frames, fridgeons, sleeping dogs and the music of the spheres. In Z. Pylyshyn, editor, *The Robot's Dilemma*, pages 139–149. Ablex, 1987.

[28] Norman Y. Foo. Observations on agm entrenchment. Technical Report 389. Department of Computer Science, University of Sydney, 1990.

[29] Michael Freund and Daniel Lehmann. Belief revision and rational inference. Technical Report TR-ARP-94-16, The Leibniz Center for Research in Computer Science, Institute of Computer Science, Hebrew University, July 1994. 1994.

[30] Nir Friedman and Joseph Halpern. Belief revision: A critique. In *Proceedings of Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, 1996.

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

[32] Andre Fuhrmann and Sven Ove Hansson. A survey of multiple contractions. *Journal of Logic, Language and Information*, 3:39–76, 1994.

[33] Peter Gärdenfors. Belief revision and irrelevance. *Philosophy of Science Association Journal*, 2:349–356.

[34] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Belief States.* Bradford Books, MIT Press, Cambridge, 1988.

[35] Peter Gärdenfors. Belief revision and nonmonotonic logic: Two sides of the same coin? In L. Carlucci Aiello, editor, *Proceedings of the Ninth European Conference on Artificial Intelligence*, pages 768–773, London, 1990. Pitman Publishing.

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

[37] Peter Gärdenfors and Hans Rott. Belief revision. In Dov M. Gabbay, Christopher J. Hogger, and John A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume IV: Epistemic and Temporal Reasoning*, pages 35–132. Oxford University Press, Cambridge, 1995.

[38] Konstantinos Georgatos. Ordering-based representations of rational inference. In J.J. Alferes, L.M. Pereira, and E. Orlowska, editors, *Logics in Artificial Intelligence (JELIA '96)*. Lecture Notes in Artificial Intelligence, pages 176–191, Berlin, 1996. Springer-Verlag.

[39] Konstantinos Georgatos. Entrenchment relations: A uniform approach to nonmonotonicity. In *Proceedings of the International Joint Conference on Qualitative and Quantitative Practical Reasoning (ESCQARU/FAPR 97)*, Lecture Notes in Computer Science, Number 1244, pages 282–297, Berlin, 1997. Springer-Verlag.

[40] Matthew Ginsberg. Multivalued logics: A uniform approach to reasoning in artificial intelligence. *Computational Intelligence*, 4:265–316, 1988.

[41] Matthew Ginsberg. Do computers need common sense? In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR'96)*, California, 1996. Morgan Kaufmann.

[42] M. Goldzmit and Judea Pearl. Rank based systems: A simple approach to belief revision, belief update and reasoning about evidence and other actions.

In *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference (KR '92)*, pages 661–672, 1992.

[43] Nelson Goodman. About. *Mind*, 70:1–24, 1961.

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

[45] Sven Ove Hansson. New operators for theory change. *Theoria*, 35:114–132, 1989.

[46] Sven Ove Hansson. Belief contraction without recovery. *Studia Logica*. 50:250–260, 1991.

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

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

[49] Sven Ove Hansson. Reversing the levi identity. *Journal of Philosophical Logic*, 22:637–669, 1993.

[50] Sven Ove Hansson. Kernel contraction. *Journal of Symbolic Logic*, 59:845–859. 1994.

[51] Sven Ove Hansson. A survey of non-prioritized belief revision. *Erkenntnis*. 1999. To appear.

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

[53] Gilbert Harman. *Change in View*. MIT Press, Cambridge, 1986.

[54] W.L Harper. Rational conceptual change. In F. Suppe and P.D Asquith, editors, *PSA 1976 Vol. 2*, pages 462–494. Philosophy of Science Association, East Lansing, Michigan, 1977.

[55] Andreas Herzig and Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence*, 1999.

[56] Jakko Hintikka. Impossible worlds vindicated. *Journal of Philosophical Logic*. 4:475–484, 1975.

[57] Stanislaw Jaskowski. Propositional calculus for contradictory deductive systems. *Studia Logic*, 24:143–157, 1969.

[58] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[59] Hirofumi Katsuno and Alberto O. Mendelzon. A unified view of propositional knowledge base updates. In *Proceedings of 11th International Joint Conference on Artificial Intelligence*, pages 1413–1419, 1991.

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

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

[62] Henry Kyburg. The rule of adjunction and reasonable inference. *Journal of Philosophy*, XCIV(3):109–125, March 1997.

[63] Daniel Lehmann. Belief revision, revised. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1534–1540, 1995.

[64] Isaac Levi. Subjunctives, dispositions, chances. *Synthese*, 34:423–455, 1977.

[65] Isaac Levi. *The Enterprise of Knowledge*. MIT Press, 1983.

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

[67] David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973.

[68] David Lewis. Logic for equivocators. In *Papers in Philosophical Logic*. Cambridge University Press, 1998.

[69] David Makinson. How to give it up: A survey of some formal aspects of the logic of theory change. *Synthese*, 62:347–363, 1985.

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

[71] James McLelland. Constructive memory and memory distortions: A parallel-distributed processing approach. In Daniel Schacter, editor, *Memory Distortion*. Harvard University Press, 1995.

[72] A. C. Nayak, N.Y. Foo, M. Pagnucco, and A. Sattar. Changing conditional beliefs unconditionally. In *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, pages 119–135. De Zeeuwse Stromen. The Netherlands, 1996. Morgan Kaufmann.

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

[74] Bernhard Nebel. *Reasoning and revision in hybrid representation systems, volume 422 of Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1990.

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

[76] Bernhard Nebel. A knowledge level analysis of belief revision. In R. Brachman, Hector Levesque, and Raymond Reiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*. pages 301–311, San Mateo, CA, 1996. Morgan-Kaufmann.

[77] Bernhard Nebel. How hard is it to revise a belief base? In D. Dubois and H. Prade, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3: Belief Change*, pages 77–145. Kluwer Academic, Dordrecht, The Netherlands, 1998.

[78] R.E Nisbett and L.D Ross. *Human Inference: Strategy and Shortcomings*. Prentice Hall, Englewood-Cliffs, 1980.

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

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

[81] Michael Pendlebury. Stalnaker on inquiry. *Journal of Philosophical Logic*, 16:229–272, 1987.

[82] Graham Priest. Paraconsistent logic. In Dov Gabbay, editor, *Handbook of Philosophical Logic*.

[83] Graham Priest. Paraconsistent belief revision. *Theoria*, 1999. To appear.

[84] Graham Priest and R. Routley. Systems of paraconsistent logic. In Graham Priest, R. Routley, and J. Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*. Philosophia Verlag, 1989.

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

[86] W.V.O Quine. Two dogmas of empiricism. In *From a Logical Point of View*. Cambridge University Press, Cambridge, 1953.

[87] V. Rantala. Impossible world semantics and logical omniscience. *Acta Philosophica Fennica*, 35:18–24, 1982.

[88] Nicholas Rescher. *Hypothetical Reasoning*. North Holland, Amsterdam, 1964.

[89] Nicholas Rescher and Robert Brandom. *The Logic Of Inconsistency: A Study in Non-Standard Possible-World Semantics and Ontology*. Rowman and Littlefield, 1978.

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

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

[92] Hans Rott. Two methods of constructing contractions and revisions of knowledge systems. *Journal of Philosophical Logic*, 20:149–173, 1991.

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

[94] Hans Rott. *Change, Choice and Inference*. Oxford University Press, 1996.

[95] Mark D. Ryan. Ordered theory presentations. In *Proceedings of the 8th Amsterdam Colloquium*, 1991.

[96] Karl Schlechta. Some results on theory revision. In Andre Fuhrmann and N. Morreau, editors, *The Logic of Theory Change*, Lecture Notes in Artificial Intelligence, Number 465. Springer, Berlin, 1991.

[97] Schotch and Jennings. On detonating. In Graham Priest, R. Routley, and J. Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*. Philosophia Verlag, 1989.

[98] P.K Schotch and R.E Jennings. Inference and necessity. *Journal of Philosophical Logic*, 9:327–240, 1980.

[99] Robert Stalnaker. *Inquiry*. MIT Press, 1986.

[100] Stephen P. Stich. *The Fragmentation of Reason*. MIT Press, 1991.

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

[102] Moshe Vardi. The complexity of epistemic reasoning. In *Sixth Annual IEEE Symposium on Logic in Computer Science*, 1989.

[103] P. Wason and P. Johnson-Laird. *Psychology of Reasoning: Structure and Content*. B.T Batsford, London, 1972.

[104] Renata Wasserman. Full acceptance through argumentation. In *Proceedings of IJCAI-99 Workshop on Practical Reasoning and Rationality*, 1999.

[105] Renata Wasserman. Resource bounded belief revision. *Erkenntnis*, 1999. To appear.

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

[107] Bernard Williams. Deciding to believe. In H. Kiefer and M.K Munitz, editors, *Language, Belief and Metaphysics*, pages 95–111. SUNY Press, Albany, 1970.

[108] Timothy Williamson. *Vagueness*. Routledge, London, 1994.

[109] Marianne Winslett. Reasoning about action using a possible models approach. In *Proceedings of the 7th National Conference of the American Association for Artificial Intelligence*, pages 89–93, 1988.