Distance Semantics for Relevance-Sensitive Belief Revision

Pavlos Peppas
Dept of Business Administration
University of Patras
265 00 Patras, Greece
ppeppas@otenet.gr

Samir Chopra
Dept of Computer and
Information Science
Brooklyn College of the
City University of New York
Brooklyn, NY 11210, USA
schopra@sci.brooklyn.cuny.edu

Norman Foo
Knowledge Systems Group
School of Computer Science
and Engineering
University of New South Wales
Sydney, NSW 2052, Australia
norman@cse.unsw.edu.au

Abstract

Possible-world semantics are provided for Parikh's relevance-sensitive model for belief revision. Having Grove's system-of-spheres construction as a base, we consider additional constraints on measuring distance between possible worlds, and we prove that, in the presence of the AGM postulates, these constraints characterize precisely Parikh's axiom (P). These additional constraints essentially generalize a criterion of similarity that predates axiom (P) and was originally introduced in the context of Reasoning about Action. A by-product of our study is the identification of two possible readings of Parikh's axiom (P), which we call the *strong* and the *weak* versions of the axiom. An interesting feature of the strong version is that, unlike classical AGM belief revision, it makes associations between the revision policies of *different* theories.

Introduction

Much of the work in the field of Belief Revision is based on the classical work of Alchourron, Gardenfors and Makinson, (Alchourron, Gardernfors, & Makinson 1985), that has given rise to a formal framework for studying this process, commonly referred to as the *AGM paradigm*. Within the AGM paradigm there are two constituents that are of particular interest for this paper. The first is the set of rationality postulates for belief revision, known as the *AGM postulates* (Alchourron, Gardernfors, & Makinson 1985). The second is a special kind of preorder on possible worlds, called a *system of spheres*, based on which Grove defined a constructive model for belief revision; Grove has shown in (Grove 1988) that the AGM postulates are sound and complete with respect to his system-of-spheres semantics.

Studying the AGM paradigm, Parikh (Parikh 1999) observed that is it rather liberal in its treatment of the notion of *relevance*. More precisely, Parikh argues that during belief revision a rational agent does not change her entire belief corpus, but only the portion of it that is relevant to the new information. This intuition of *local change*, Parikh claims, is not fully captured by the AGM paradigm. To remedy this shortcoming, Parikh introduced an additional axiom, named (P), as a supplement to the AGM postulates. Loosely speaking, axiom (P) says that when new information φ is received, only part of the initial belief set K will be affected; namely the part that shares common propositional variables with the minimal language of φ. Parikh's approach is also known as the *language splitting model*.

Although axiom (P) is just a first step towards capturing the role of relevance in belief revision[1], Parikh's work has received considerable attention since the publication of (Parikh 1999) (see for example (Chopra & Parikh 1999), (Chopra & Parikh 2000), (Chopra, Georgatos, & Parikh 2001)). Yet, despite all the research on axiom (P), no semantics for it have yet been formulated. This is the gap that the present article aims to fill. We examine new constraints on systems-of-spheres and, building on Grove's result, we prove that in the presence of the AGM postulates, axiom (P) is sound and complete with respect to these new semantic constraints.

What is particularly pleasing about our result is that the new constraints on systems of spheres are in fact not new at all; they essentially generalize a very natural condition that predates axiom (P) and has been motivated independently by Winslett in the context of Reasoning about Action (Winslett 1988). This connection between Belief Revision with Reasoning about Action further confirms intuitions about the relationship between the two areas (see (Peppas & Wobcke 1992), (Peppas 1994), and (Peppas, Foo, & Nayak 2000)).

In the course of formulating semantics for axiom (P) we observed that there are in fact two possible readings of this axiom, which we call the *strong* and the *weak* versions of (P). We present both these versions herein and we show that the strong version of (P) brings with it a new feature in the picture of classical AGM revision: it makes associations between the revision policies of *different* theories.

The outline of the paper is as follows. We first present some background material on the AGM paradigm and the

[1] Work on relevance in general in Artificial Intelligence has been ongoing for a while. A comprehensive collection of papers dealing with relevance could be found in (AIJ 1997).

language splitting model (first three sections). The crucial axiom (P) is then examined in greater detail to fully flesh out its possible readings (4th section). We proceed with the formulation of semantics for axiom (P). For ease of exposition and clarity, we start by focusing on the special case of "opinionated" agents, that is, agents whose belief set is a consistent complete theory (5th section). Then we consider the general case of incomplete theories (6th section). The last section contains some concluding remarks.

Formal Preliminaries

Throughout this paper we work with a finite set of propositional variables $P = \{p_1, \dots p_m\}$. We define $\mathcal{L}$ to be the propositional language generated from P (using the standard boolean connectives $\wedge, \vee, \rightarrow, \neg$ and the special symbols $\top, \bot$) and governed by classical propositional logic $\vdash$. A sentence $\varphi \in \mathcal{L}$ is *contingent* iff $\not\vdash \varphi$ and $\not\vdash \neg\varphi$. For a set of sentences Γ of $\mathcal{L}$, we denote by $Cn(\Gamma)$ the set of all logical consequences of Γ, i.e., $Cn(\Gamma) = \{\varphi \in \mathcal{L}\colon \Gamma \vdash \varphi\}$. We shall often write $Cn(\varphi_1, \varphi_2, \dots, \varphi_n)$, for sentences φ_1, $\varphi_2, \dots, \varphi_n$, as an abbreviation of $Cn(\{\varphi_1, \varphi_2, \dots, \varphi_n\})$.

A theory T of $\mathcal{L}$ is any set of sentences of $\mathcal{L}$ closed under $\vdash$, i.e., $T = Cn(T)$. In this paper we focus only on *consistent* theories. Hence from now on, whenever the term "theory" appears unqualified, it is understood that it refers to a consistent theory. We denote the set of all consistent theories of $\mathcal{L}$ by $\mathcal{K}_{\mathcal{L}}$. A theory T of $\mathcal{L}$ is *complete* iff for all sentences $\varphi \in \mathcal{L}$, $\varphi \in T$ or $\neg\varphi \in T$. We denote the set of all consistent complete theories of $\mathcal{L}$ by $\mathcal{M}_{\mathcal{L}}$. In the context of systems of spheres, consistent complete theories essentially play the role of possible worlds. Following this convention, in the rest of the article we use the terms "possible world" (or simply "world") and "consistent complete theory" interchangeably. For a set of sentences Γ of $\mathcal{L}$, $[\Gamma]$ denotes the set of all consistent complete theories of $\mathcal{L}$ that contain Γ. Often we use the notation $[\varphi]$ for a sentence $\varphi \in \mathcal{L}$, as an abbreviation of $[\{\varphi\}]$. For a theory T and a set of sentences Γ of $\mathcal{L}$, we denote by $T + \Gamma$ the closure under $\vdash$ of $T \cup \Gamma$, i.e., $T + \Gamma = Cn(T \cup \Gamma)$. For a sentence $\varphi \in \mathcal{L}$ we often write $T + \varphi$ as an abbreviation of $T + \{\varphi\}$.

In the course of this paper, we often consider *sublanguages* of $\mathcal{L}$. Let P' be a subset of the set of propositional variables P . By $\mathcal{L}^{P'}$ we denote the sublanguage of $\mathcal{L}$ defined over P'. In the limiting case where P' is empty, we take $\mathcal{L}^{P'}$ to be the language generated by $\top, \bot$ and the boolean connectives. For a sublanguage $\mathcal{L}'$ of $\mathcal{L}$ defined over a subset P' of P, by $\overline{\mathcal{L}'}$ we denote the sublanguage defined over the propositional variables in the complement of P' i.e., $\overline{\mathcal{L}'} = \mathcal{L}^{(P-P')}$. For a sentence χ of $\mathcal{L}$, by $\mathcal{L}_\chi$ we denote the *minimal* sublanguage of $\mathcal{L}$ within which χ can be expressed (i.e., $\mathcal{L}_\chi$ contains a sentence that is logically equivalent to χ, and moreover no *proper* sublanguage of $\mathcal{L}_\chi$ contains such a sentence).[2] Finally we note that in the forthcoming discussion, we often project operations defined earlier for the entire language $\mathcal{L}$, to one of its sublanguages $\mathcal{L}'$. When this happens, all notation will be subscripted by the sublanguage $\mathcal{L}'$. For example, for a set of sentences $\Gamma \subset \mathcal{L}'$, the term $Cn_{\mathcal{L}'}(\Gamma)$ denotes the logical closure of Γ *in* $\mathcal{L}'$. Similarly, $[\Gamma]_{\mathcal{L}'}$ denotes the set of all maximally consistent supersets of Γ *in* $\mathcal{L}'$. When no subscript is present, it is understood that the operation is relevant to the original language $\mathcal{L}$.

[2] It is not hard to verify that for every χ, $\mathcal{L}_\chi$ is unique – see (Parikh 1999) for details.

The AGM Paradigm

Much research in belief revision is based on the work of Alchourron, Gardenfors and Makinson (Alchourron, Gardernfors, & Makinson 1985), who have developed a research framework for this process, known as the *AGM paradigm*. In this section we shall briefly review two of the main models for belief revision within the AGM paradigm; the first is based on a set of *rationality postulates*, and the second on a preorder on worlds known as a *system of spheres*.

The AGM Postulates

In the AGM paradigm, belief revision is modeled as a function $*$ mapping a theory T and a sentence φ to the theory $T * \varphi$. In this paper we assume that the theory T and the sentence φ are *individually* consistent. Alchourron, Gardenfors, and Makinson have proposed the following set of postulates for belief revision:

(K*1) $T * \varphi$ is a theory of $\mathcal{L}$.

(K*2) $\varphi \in T * \varphi$.

(K*3) $T * \varphi \subseteq T + \varphi$.

(K*4) If $\neg\varphi \notin T$ then $T + \varphi \subseteq T * \varphi$.

(K*5) $T * \varphi = \mathcal{L}$ iff $\vdash \neg\varphi$.

(K*6) If $\vdash \varphi \leftrightarrow \psi$ then $T * \varphi = T * \psi$.

(K*7) $T * (\varphi \wedge \psi) \subseteq (T * \varphi) + \psi$.

(K*8) If $\neg\psi \notin T * \varphi$ then $(T * \varphi) + \psi \subseteq T * (\varphi \wedge \psi)$.

Systems of Spheres

Apart from axiomatic approaches to belief revision, a number of explicit constructions for this process have been proposed. One popular construction is that proposed by Grove (Grove 1988) based on a total preorder on possible worlds.

Definition 1 *(Grove 1988) Let T be a theory of $\mathcal{L}$, and S_T a collection of sets of possible worlds i.e., $S_T \subseteq 2^{\mathcal{M}_{\mathcal{L}}}$. S_T is a* system of spheres centered on $[T]$ *iff the following conditions are satisfied:*[3]

[3] We include condition (S4) for reasons of comprehensiveness, even though in the finite propositional case, this condition is redundant.

(S1) *S_T is totally ordered with respect to set inclusion; that is, if $V, U \in S_T$ then $V \subseteq U$ or $U \subseteq V$.*

(S2) *The smallest sphere in S_T is $[T]$; that is, $[T] \in S_T$, and if $V \in S_T$ then $[T] \subseteq V$.*

(S3) *$\mathcal{M}_\mathcal{L} \in S_T$ (and therefore $\mathcal{M}_\mathcal{L}$ is the largest sphere in S_T).*

(S4) *For every $\varphi \in \mathcal{L}$, if there is any sphere in S_T intersecting $[\varphi]$ then there is also a smallest sphere in S_T intersecting $[\varphi]$.*

For a system of spheres S_T and a sentence $\varphi \in \mathcal{L}$, the smallest sphere in S_T intersecting $[\varphi]$ is denoted $C_T(\varphi)$.[4] With any system of spheres S_T, Grove associates a function $f_T : \mathcal{L} \mapsto 2^{\mathcal{M}_\mathcal{L}}$ defined as follows:

$$f_T(\varphi) = [\varphi] \cap C_T(\varphi)$$

Grove uses the system of spheres S and its associated function f_T, to define constructively the process of revising T, by means of the following condition:

(S*) $T * \varphi = \bigcap f_T(\varphi)$

Grove proved that the class of functions generated from systems of spheres by means of (S*) is precisely the family of revision functions satisfying the eight AGM postulates (K*1) - (K*8). One of the main aims of this paper is to characterize the subclass of systems of spheres which correspond (via (S*)) to revision functions that, in addition to (K*1) - (K*8), also satisfy axiom (P) (see below).

Relevance-Sensitive Belief Revision

When revising a theory T by a sentence φ it seems plausible to assume that only the beliefs that are *relevant* to φ should be affected, while the rest of the belief corpus is unchanged. For example, an agent that is revising her beliefs about planetary motion is unlikely to revise her beliefs about Malaysian politics. This simple intuition is not fully captured in the AGM paradigm. To see this consider the trivial revision function $*_t$ defined below:

$$T *_t \varphi = \begin{cases} T + \varphi & \text{if } \varphi \text{ is consistent with } T \\ Cn(\varphi) & \text{otherwise} \end{cases}$$

It is not hard to verify that $*_t$ satisfies all the AGM postulates, and yet it has the rather counter-intuitive effect of throwing away *all* non-tautological beliefs in T whenever the new information φ is inconsistent with T, regardless of whether these beliefs are related to φ or not.

[4]In the limiting case where φ is inconsistent, Grove defines $C_T(\varphi)$ to be the set $\mathcal{M}_\mathcal{L}$. However, since in this paper we only consider revision by *consistent* sentences, these limiting cases are irrelevant.

In order to block revision functions like $*_t$ Parikh introduced in (Parikh 1999) a new axiom, named (P), as a supplement to the AGM postulates. The main intuition that axiom (P) aims to capture is that an agent's beliefs can be subdivided into disjoint compartments, referring to different subject matters, and that when revising, the agent modifies only the compartment(s) affected by the new information:

(P) If $T = Cn(\chi, \psi)$ where χ, ψ are sentences of disjoint sublanguages $\mathcal{L}_1$, $\mathcal{L}_2$ respectively, and $\varphi \in \mathcal{L}_1$, then $T * \varphi = (Cn_{\mathcal{L}_1}(\chi) \circ \varphi) + \psi$, where $\circ$ is a revision operator of the sublanguage $\mathcal{L}_1$.

It was shown in (Parikh 1999) that (P) is consistent with the AGM postulates, (K*1) - (K*6) (known as the *basic* AGM postulates). The results presented later in this paper entail that (P) is in fact consistent with all eight AGM postulates (K*1) - (K*8).

Two Readings of Axiom (P)

Before proceeding with the formulation of semantics for axiom (P), it is worth taking a closer look at it.

Consider two sentences $\chi, \psi \in \mathcal{L}$, such that $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$, and let T be the theory $T = Cn(\{\chi, \psi\})$. Moreover, let φ be any sentence in $\mathcal{L}_\chi$. According to axiom (P), anything outside $\mathcal{L}_\chi$ will not be affected by the revision of T by φ. This however is only one side of axiom (P). The other side concerns the part of the theory T that *is* related to φ, which according to axiom (P) will change to $Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi$, where $\circ$ is a revision function defined over the sublanguage $\mathcal{L}_\chi$. It is this second side of axiom (P) that needs closer examination.

Axiom (P) is open to two different interpretations. According to the first reading, which we call the *weak* version of axiom (P), the revision function $\circ$ that modifies the relevant part of T – call it the *local* revision function – may *vary* from theory to theory, even when the relevant part $Cn(\chi)$ stays the same. To give a concrete example, let a, b, c be propositional variables, let T be the theory $T = Cn(a \wedge b, c)$, and let T' be the theory $T' = Cn(a \wedge b, \neg c)$. Denote by $\mathcal{L}_1$ the sublanguage defined over $\{a, b\}$ and by $\mathcal{L}_2$ the sublanguage defined over $\{c\}$. Moreover, let φ be the sentence $\varphi = \neg a \vee \neg b$. The part of T and T' that is relevant to φ (in the sense of the language-splitting model) is the same for both theories, namely $Cn(a \wedge b)$. Nevertheless, according to the *weak* version of axiom (P), the local revision operators $\circ$ and $\circ'$ that modify the two *identical* relevant parts of T and T' respectively, may very well *differ*. For example, it could be the case that $Cn_{\mathcal{L}_1}(a \wedge b) \circ (\neg a \vee \neg b) = Cn_{\mathcal{L}_1}(\neg a \wedge b)$, and $Cn_{\mathcal{L}_1}(a \wedge b) \circ' (\neg a \vee \neg b) = Cn_{\mathcal{L}_1}(a \wedge \neg b)$, from which it follows that $T * \varphi = Cn(\neg a, b, c)$, and $T' * \varphi = Cn(a, \neg b, \neg c)$. In other words, the weak version of axiom (P) allows the local revision function to be *context-sensitive*. In the scenario described above, the presence of c in T leads to a local revision function $\circ$ for $Cn_{\mathcal{L}_1}(a \wedge b)$ that produces $Cn_{\mathcal{L}_1}(\neg a \wedge b)$ as the result of revising by $\neg a \vee \neg b$; on the other hand, the presence of $\neg c$ in T', induces a local revision function $\circ'$

for $Cn_{\mathcal{L}_1}(a \wedge b)$ that produces $Cn_{\mathcal{L}_1}(a \wedge \neg b)$ for the same input. Therefore, while c (or $\neg c$) remains unaffected during the (global) revision by $\neg a \vee \neg b$ (since it is not relevant to the new information), its presence influences the way that the relevant part of the theory is modified.

To prevent such an influence we need to resort to the *strong* version of axiom (P) which makes the local revision function $\circ$ *context-independent*. According to the strong interpretation of (P), for any two theories $T = Cn(\chi, \psi)$ and $T' = Cn(\chi, \psi')$, such that $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \mathcal{L}_\chi \cap \mathcal{L}_{\psi'} = \emptyset$, there exists a *single* local revision function $\circ$ such that $T * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi$ and $T' * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi'$, for any $\varphi \in \mathcal{L}_\chi$.

It should be noted that although axiom (P) is open to both the weak and the strong interpretations, the discussion and some results in (Parikh 1999) suggest that the strong version of axiom (P) is intended. Following Parikh, we shall also adopt the strong version of axiom (P) in this paper. To make this assumption explicit and to avoid any ambiguity, we make use of the following two conditions which together are shown to be equivalent to the strong version of axiom (P):

(R1) If $T = Cn(\chi, \psi)$, $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$, and $\varphi \in \mathcal{L}_\chi$, then $(T * \varphi) \cap \overline{\mathcal{L}_\chi} = T \cap \overline{\mathcal{L}_\chi}$.

(R2) If $T = Cn(\chi, \psi)$, $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$, and $\varphi \in \mathcal{L}_\chi$, then $(T * \varphi) \cap \mathcal{L}_\chi = (Cn(\chi) * \varphi) \cap \mathcal{L}_\chi$.

Condition (R1) is straightforward: when revising a theory T by a sentence φ, the part of T that is *not related* to φ is not affected by the revision. Condition (R2) is what imposes the strong version of axiom (P). To see this, consider a revision function $*$ (which defines a revision policy for *all* the theories of $\mathcal{L}$), and let $T = Cn(\chi, \psi)$ and $T' = Cn(\chi, \psi')$ be two theories such that $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \mathcal{L}_\chi \cap \mathcal{L}_{\psi'} = \emptyset$. Consider now any sentence $\varphi \in \mathcal{L}_\chi$. The relevant part to φ of T and T' is in both cases the same. Then, according to (R2), the way that this relevant part is modified in both T and T' is also the same; namely, as dictated by the revision function $*$ itself when applied to $Cn(\chi)$ (once again, notice that $*$ is defined for all theories, including T, T', *and* $Cn(\chi)$).

The following result shows that (R1) and (R2) are indeed equivalent with the strong version of axiom (P).

Theorem 1 *Let $*$ be a revision function satisfying the AGM postulates (K*1) - (K*8). Then $*$ satisfies (P) iff $*$ satisfies (R1) and (R2).*

Proof.

$(\Rightarrow)$

Assume that $*$ satisfies (P). Let T be a theory of $\mathcal{L}$ such that $T = Cn(\chi, \psi)$, where $\chi, \psi \in \mathcal{L}$ and $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$. Consider now any sentence $\varphi \in \mathcal{L}_\chi$. By (P) it follows that $T * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi$, where $\circ$ is a revision operator of $\mathcal{L}_\chi$. From the above, (R1) follows immediately. For (R2), let us denote by T' the theory $Cn(\chi)$. Firstly notice that $T' * \varphi$ is equal to the closure in $\mathcal{L}$ of $Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi$ i.e., $T' * \varphi = Cn(Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi)$. Indeed, T' can be written as $T' = Cn(\chi, \psi \vee \neg \psi)$ and therefore, by (the strong version of) axiom (P), $T' * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + (\psi \vee \neg \psi) = Cn(Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi)$. Consequently, $T * \varphi = (Cn(\chi) * \varphi) + \psi$, and moreover $\mathcal{L}_{Cn(\chi)*\varphi} \cap \mathcal{L}_\psi = \emptyset$. Therefore, $(T * \varphi) \cap \mathcal{L}_\chi = (Cn(\chi) * \varphi) \cap \mathcal{L}_\chi$ as desired.

$(\Leftarrow)$

Assume that $*$ satisfies (R1) and (R2). Let T be a theory of $\mathcal{L}$ such that $T = Cn(\chi, \psi)$, where $\chi, \psi \in \mathcal{L}$ and $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$. As a first step in proving (P), we shall show that for any $\varphi \in \mathcal{L}_\chi$, the theory $T * \varphi$ is also split between $\mathcal{L}_\chi$ and $\mathcal{L}_\psi$. Assume on the contrary that this is not the case for a particular $\varphi \in \mathcal{L}_\chi$. Then, as shown by Lemma-A in (Parikh 1999), there is a world r such that both $r \cap \mathcal{L}_\chi$ and $r \cap \overline{\mathcal{L}_\chi}$ are *individually* consistent with $T * \varphi$, and yet $r \notin [T * \varphi]$. Let α be the conjunction of literals in $\mathcal{L}_\chi$ that hold at r, and similarly, let β be the conjunction of literals in $\overline{\mathcal{L}_\chi}$ that hold at r. Since α is consistent with $T * \varphi$, from (K*7) and (K*8) it follows that $T * (\varphi \wedge \alpha) = (T * \varphi) + \alpha$. Therefore, $r \notin [T * (\varphi \wedge \alpha)]$. Consider now any world r' in $[T * (\varphi \wedge \alpha)]$. Clearly, $r' \vdash \alpha$, and given that $r' \neq r$, it follows that $r' \vdash \neg \beta$. Since all worlds in $[T * (\varphi \wedge \alpha)]$ satisfy $\neg \beta$ it follows that $\neg \beta \in T * (\varphi \wedge \alpha)$. Then, by (R1) we derive that $\neg \beta \in T$, which again by (R1) entails that $\neg \beta \in T * \varphi$. This however contradicts the fact that $r \cap \overline{\mathcal{L}_\chi}$ is consistent with $T * \varphi$. Hence we have shown that, given (R1), the theory $T * \varphi$ is split between $\mathcal{L}_\chi$ and $\mathcal{L}_\psi$ for all $\varphi \in \mathcal{L}_\chi$.

Continuing with the proof of condition (P), we define $\circ$ to be the following operator of $\mathcal{L}_\chi$: for any theory T' of $\mathcal{L}_\chi$ and any sentence $\gamma \in \mathcal{L}_\chi$, $T' \circ \gamma = (Cn(T') * \gamma) \cap \mathcal{L}_\chi$. It is not hard to verify that $\circ$ is indeed an AGM revision operator, i.e., it satisfies the postulates (K*1) - (K*8). Consider now any sentence $\varphi \in \mathcal{L}_\chi$. We conclude the proof of this theorem by showing that $T * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi$.

Indeed, from (R2) it follows that $(T * \varphi) \cap \mathcal{L}_\chi = (Cn(\chi) * \varphi) \cap \mathcal{L}_\chi$, and therefore by the construction of $\circ$, $(T * \varphi) \cap \mathcal{L}_\chi = Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi$. Moreover, since $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$, we derive that $(T * \varphi) \cap \mathcal{L}_\chi = ((Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi) \cap \mathcal{L}_\chi$. On the other hand, from (R1) we have that $(T * \varphi) \cap \overline{\mathcal{L}_\chi} = T \cap \overline{\mathcal{L}_\chi} = Cn(\chi, \psi) \cap \overline{\mathcal{L}_\chi} = ((Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi) \cap \overline{\mathcal{L}_\chi}$ (the last equation follows from the fact that $(Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi)$ is in $\mathcal{L}_\chi$ which in turn is disjoint from $\mathcal{L}_\psi$). Putting together the above observations we notice that the two theories, $T * \varphi$ and $(Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi$ are identical when projected on $\mathcal{L}_\chi$ *as well as* when projected on to its complement $\overline{\mathcal{L}_\chi}$. Given that, as shown earlier, $T * \varphi$ is split between $\mathcal{L}_\chi$ and $\mathcal{L}_\psi$, we derive the desired identity; i.e. $T * \varphi = (Cn_{\mathcal{L}_\chi}(\chi) \circ \varphi) + \psi$[5]
■

The strong version of axiom (P) brings about a new feature in the picture of classical AGM revision: it makes asso-

[5]Notice the use of (R2) in proving the strong version of axiom (P).

ciations between the revision policies of *different* theories. None of the AGM postulates have this property – they all refer to a *single* theory T – making any combination of revision policies on different theories permissible (as long of course as each policy *individually* satisfies the AGM axioms). This is no longer the case when (R2) (or the strong version of (P)) is brought into the picture. This condition introduces dependencies between the revisions carried out on different (overlapping) theories.

The Special Case of Complete Theories

Let us now turn to our main objective in this article, which is to formulate system-of-spheres semantics for axiom (P). The semantics will be developed progressively in two steps. In the first step, undertaken in this section, we limit ourselves to the first side of axiom (P), i.e., condition (R1). Moreover we consider only *complete* theories as belief sets. Then, in the second step (next section), we generalize our results to arbitrary theories and we also bring (R2) into the picture.

The reason for this two-phase approach is mainly to increase readability and enhance the motivation of the concepts that will be introduced later. Conditions (R1) and (R2) are quite independent of one another so it makes sense to study them separately. Moreover, the characterization of (R1) in terms of systems of spheres is much more intuitive when confined to complete theories; once this characterization is well understood for the special case, its generalization, although not trivial, is easier to follow.

Let T be a consistent complete theory, and let S_T be a system of spheres centered on $[T]$. The intended reading of S_T is that it represents *comparative similarity* between possible worlds i.e., the further away a world is from the center of S_T, the less similar it is to $[T]$.[6] None of the conditions (S1) - (S4) however indicate *how* similarity between worlds should be measured. In (Peppas, Foo, & Nayak 2000) a specific criterion of similarity is considered, originally introduced in the context of *Reasoning about Action* with Winslett's *Possible Models Approach* (PMA) (Winslett 1988). This criterion, called *PMA's criterion of similarity*, measures "distance" between worlds based on propositional variables. In particular, let r, r' be any two possible worlds of $\mathcal{L}$. By *Diff*(r, r') we denote the set of propositional variables that have different truth values in the two worlds i.e., *Diff*(r, r') = $\{p_i \in P : p_i \in r$ and $p_i \notin r'\} \cup \{p_j \in P : p_j \notin r$ and $p_j \in r'\}$. A system of spheres S_T is a *PMA system of spheres* iff it satisfies the following condition (Peppas, Foo, & Nayak 2000) (throughout this paper, the symbols r and r' always represent consistent complete theories):

(PS) If *Diff*$(T, r) \subset$ *Diff*(T, r') then there is a sphere $V \in S_T$ that contains r but not r'.

[6] Perhaps "comparative plausibility" would have been a better term in the present context. However we shall tolerate this slight abuse of terminology mainly to comply with (Peppas, Foo, & Nayak 2000).

According to condition (PS), the less a world r differs from the initial belief set T in propositional variables, the closer it is to the center of S_T. Notice that condition (PS) places no constraints on the relative order of worlds that are *Diff*-incomparable. In other words, for two worlds r and r' such that neither *Diff*$(T, r) \subset$ *Diff*(T, r') nor *Diff*$(T, r') \subset$ *Diff*(T, r), their relative order in S_T is not constrained by (PS).

It turns out that, in the special case of consistent complete belief sets, condition (PS) is the counterpart of (R1) in the realm of systems of spheres. Before however presenting the formal result, let us consider intuitively why this might be so.

Let S_T be a system of spheres centered on $[T]$ that satisfies (PS). Moreover let φ be any consistent sentence that contradicts T (i.e., $\neg\varphi \in T$). The set of φ-worlds occupy a territory in S_T that is disjoint from the center $[T]$. At the outskirts of this φ-territory there are worlds that look very different from T. However, as we move closer to the center of S_T the φ-worlds that we meet agree with T in progressively more and more propositional variables. By the time we reach the boundary of the φ-territory with the center of S_T, all the φ-worlds there agree with T in *every* propositional variable outside $\mathcal{L}_\varphi$. Hence, the intersection of these worlds (which by (S*) is the revision of T by φ) also agrees entirely with T outside $\mathcal{L}_\varphi$; thus (R1).

The above intuitive explanation of the relationship between (PS) and (R1) is formally established with following result:

Theorem 2 *Let $*$ be a revision function satisfying (K*1) - (K*8), T a consistent complete theory of $\mathcal{L}$, and S_T the system of spheres centered on $[T]$, corresponding to $*$ by means of (S*). Then $*$ satisfies (R1) at T iff S_T satisfies (PS).*

Proof.

$(\Rightarrow)$

Assume that $*$ satisfies (R1) at T. Moreover assume that, contrary to the theorem, S_T violates (PS). Let V be the smallest sphere in S_T that violates (PS). That is, V is the smallest sphere in S_T that contains a world r'' for which there exists another world r' such that *Diff*$(T, r') \subset$ *Diff*(T, r''), and moreover the smallest sphere containing r', call it U, is not a subset of V (and therefore $V \subseteq U$). Let φ be the conjunction of all the literals[7] in r' than are not in T. Notice that from *Diff*$(T, r') \subset$ *Diff*(T, r''), it follows that $\varphi \in r''$. Moreover, it is easily verified that V is the smallest sphere in S_T that intersects $[\varphi]$. Indeed, assume on the contrary that a sphere smaller than V, call it V', contains a world z satisfying φ, i.e., $\varphi \in z$. Then clearly *Diff*(T, r') $\subseteq$ *Diff*(T, z), and since $r' \neq z$, it follows that *Diff*$(T, r') \subset$ *Diff*(T, z). Hence V' violates (PS), which contradicts our initial assumption that V is the smallest sphere in S_T that

[7] A literal is a propositional variable or the negation of a proposition variable.

violates (PS). Therefore V is indeed the smallest sphere in S_T intersecting $[\varphi]$. Consequently, $r'' \in [T * \varphi]$. Consider now a literal $l \in r''$, such that $l \notin T$ and $l \notin \mathcal{L}_\varphi$. Note that since *Diff*$(T, r') \subset$ *Diff*(T, r''), such a literal l indeed exists. Clearly then, $\neg l \in T$, and $\neg l \notin T * \varphi$. This however contradicts (R1) since $\neg l \notin \mathcal{L}_\varphi$, and therefore it should have remained unaffected from the revision by φ.

($\Leftarrow$)

Assume that S_T satisfies (PS), and let χ, ψ be sentences in $\mathcal{L}$, such that $T = Cn(\chi, \psi)$ and $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$. Consider now any sentence $\varphi \in \mathcal{L}_\chi$, and let r' be any world in $[T * \varphi]$ i.e., $\varphi \in r'$ and r' belongs to the smallest sphere $C_T(\varphi)$ in S_T that intersects $[\varphi]$. Firstly we show that *Diff*$(T, r') \subseteq \mathcal{L}_\chi$. Assume on the contrary that there is a literal $l \in T \cap \overline{\mathcal{L}_\chi}$, such that $l \notin r'$. Let r'' be the consistent complete theory that agrees with r' in all literals except l. Clearly then, since $\varphi \in r'$ and $l \notin \mathcal{L}_\varphi$, we derive that $\varphi \in r''$. Moreover, by the construction of r'', *Diff*$(T, r'') \subset$ *Diff*(T, r'). Consequently, by (PS), there exists a sphere V that contains r'' and does not contain r'. Therefore $V \subset C_T(\varphi)$. This leads to a contradiction since V contains a φ-world (namely, r''), and at the same time it is smaller than the smallest sphere intersecting $[\varphi]$. Hence *Diff*$(T, r') \subseteq \mathcal{L}_\chi$. This shows that all worlds r' in $[T * \varphi]$ agree with T on everything outside $\mathcal{L}_\chi$. Therefore $T \cap \overline{\mathcal{L}_\chi} = (T * \varphi) \cap \overline{\mathcal{L}_\chi}$ as desired. ■

It is worth noting that in (Peppas, Foo, & Nayak 2000), there exists a characterization of condition (PS) in terms of *epistemic entrenchments* (Gardenfors & Makinson 1988). Consequently, since by Theorem 2 (R1) is equivalent to (PS), the results in (Peppas, Foo, & Nayak 2000) can be used to provide a characterization of (R1) in terms of epistemic entrenchments.

As already mentioned in the introduction, what is quite appealing about Theorem 2 is that it characterizes (R1), not in terms of some technical non-intuitive condition, but rather by a natural constraint on similarity between possible worlds, that in fact predates (R1) and was motivated independently in a different context (Winslett 1988). Moreover, as we will show in the next section, the essence of this characterization of (R1) in terms of constraints on similarity, carries over into the general case of incomplete belief sets (albeit with some modifications).

The General Case

To elevate Theorem 2 to the general case, we first need to extend the definition of *Diff* to cover comparisons between a world r and an arbitrary, possibly *incomplete*, theory T. The generalization of *Diff* that we shall use herein takes into account the notion of a T-*splitting* introduced by Parikh in his language-splitting model (Parikh 1999).

Let T be a theory of $\mathcal{L}$ and $P_1, P_2, \ldots, P_n$ a partition of the set P of all propositional variables in $\mathcal{L}$. We say that $\{P_1, P_2, \ldots, P_n\}$ is a T-*splitting* iff there exist sentences $\varphi_1 \in \mathcal{L}^{P_1}, \varphi_2 \in \mathcal{L}^{P_2}, \ldots, \varphi_n \in \mathcal{L}^{P_n}$, such that $T = Cn(\varphi_1, \varphi_2, \ldots, \varphi_n)$. Parikh has shown in (Parikh 1999) that for every theory T there is a unique *finest* T-splitting, i.e. one which refines[8] every other T-splitting. We shall denote the finest T-splitting of T by $\mathcal{F}(T)$.

Using the notion of a finest T-splitting, we define the difference between a (possibly incomplete) theory T of $\mathcal{L}$ and a world r as follows:

Definition 2 *Let T be a consistent theory of $\mathcal{L}$ (possibly incomplete) and r a possible world. The difference Diff(T, r) between T and r is the following subset of P:*

$$\textit{Diff}(T,r) = \bigcup\{P' \in \mathcal{F}(T) : \text{for some } \varphi \in \mathcal{L}^{P'}, T \vdash \varphi \text{ and } r \vdash \neg\varphi\}$$

It is not hard to verify that in the special case of a consistent complete theory T, the above definition of *Diff* collapses to the one given in the previous section.

Notice that if T is incomplete, then for any world w compatible with T (i.e. $w \in [T]$), *Diff*$(T, w) = \emptyset$. Moreover, for any world r, *Diff*$(T, r) \subseteq \bigcup_{w \in [T]}$ *Diff*(w, r); the precise relationship between *Diff*(T, r) and *Diff*(w, r) for $w \in [T]$, is given by the following result:

Theorem 3 *Let T be a consistent theory of $\mathcal{L}$ and r a possible world. Then, Diff$(T, r) = \bigcup\{P' \in \mathcal{F}(T)$: for all $w \in [T]$, $P' \cap$ Diff$(w, r) \neq \emptyset\}$.*

Proof.

(LHS) $\subseteq$ (RHS)

Assume that $p \in$ *Diff*(T, r). Then for some $P' \in \mathcal{F}(T)$, $p \in P'$ and for some $\varphi \in \mathcal{L}^{P'}$, $T \vdash \varphi$ and $r \vdash \neg\varphi$. Let $l_1, \ldots l_n$ be the literals in $\mathcal{L}^{P'}$ that hold in r. Clearly then, $l_1 \wedge \cdots \wedge l_n \vdash \neg\varphi$, and therefore $\varphi \vdash \neg l_1 \vee \cdots \vee \neg l_n$. From this we derive that for all $w \in [T]$, since $w \vdash \varphi$, there is at least one $1 \leq k \leq n$ such that $w \vdash \neg l_k$. Consequently, for all $w \in [T]$, $P' \cap$ *Diff*$(w, r) \neq \emptyset$.

(LHS) $\supseteq$ (RHS)

Assume that for some $P' \in \mathcal{F}(T)$, $P' \cap$ *Diff*$(w, r) \neq \emptyset$ for all $w \in [T]$. Let $l_1, \ldots l_n$ be the literals in $\mathcal{L}^{P'}$ that hold in r. Then for all $w \in [T]$ there is a $1 \leq j \leq n$ such that $\neg l_j \in w$. Therefore, $T \vdash \neg l_1 \vee \cdots \vee \neg l_n$. Since $(\neg l_1 \vee \cdots \vee \neg l_n) \in \mathcal{L}^{P'}$, it follows that $P' \subseteq$ *Diff*(T, r) as desired. ■

Condition (R1) and Systems of Spheres

Having generalized *Diff*, let us re-examine condition (PS) for a system of spheres S_T related to a belief set T that is not

[8] A partition Z refines another partition Z', if for every element of Z there is a superset of it in Z'.

necessarily complete. It turns out that in this case (PS) no longer corresponds to (R1); in particular, (PS) is too strong.

To see this, consider the following counter-example. Assume that $\mathcal{L}$ is built from the propositional variables a, b, c, d, and let T be the theory $T = Cn(a \leftrightarrow b, c \leftrightarrow d)$. Clearly, $\{\{a,b\}, \{c,d\}\}$ is a T-split. Next, let S_T be the following system of spheres (represented as a total preorder on worlds):[9]

$$\begin{array}{c} abcd \\ ab\overline{c}\overline{d} \\ \overline{a}\overline{b}cd \\ \overline{a}\overline{b}\overline{c}\overline{d} \end{array} \leq \begin{array}{c} \overline{a}bcd \\ a\overline{b}cd \\ \overline{a}b\overline{c}\overline{d} \\ a\overline{b}\overline{c}\overline{d} \end{array} \leq \begin{array}{c} abc\overline{d} \\ \overline{a}\overline{b}c\overline{d} \end{array} \leq \begin{array}{c} \overline{a}\overline{b}\overline{c}d \\ ab\overline{c}d \\ \overline{a}bc\overline{d} \end{array} \leq \begin{array}{c} \overline{a}b\overline{c}d \\ a\overline{b}\overline{c}d \\ a\overline{b}c\overline{d} \end{array}$$

It is not hard to verify that the revision function $*$ induced by the above system of spheres S_T, satisfies (R1). At the same time however, S_T violates (PS). In particular consider the worlds $r = \{a, b, \neg c, d\}$ and $r' = \{\neg a, b, c, \neg d\}\}$. Although *Diff*$(T, r) = \{c, d\} \subset$ *Diff*$(T, r') = \{a, b, c, d\}$, the two worlds r and r' are equidistant from the center of S_T.

Despite its failure to generalize, (PS) should not be disregarded altogether. It can still serve as a guide in formulating the appropriate counterpart(s) of (R1) for the general case; as we prove later in this section, the two general conditions (Q1) and (Q2) that correspond to (R1) are both in the spirit of (PS) (and not surprisingly, they collapse to (PS) in the special case of complete belief sets).

To formulate the conditions (Q1) and (Q2), we first need to introduce some further concepts related to the notion of distance between a world and an incomplete theory.

Consider a theory T and let r be a world not compatible with T i.e., $r \notin [T]$. Clearly *Diff*$(T, r) \neq \emptyset$. Is there another world r' that differs from T on exactly the same propositional variables, i.e., *Diff*$(T, r) =$ *Diff*(T, r')? If T is complete, the answer is obviously "no": for any set of propositional variables P', there can only be *one* world r such that *Diff*$(T, r) = P'$. If however T is *incomplete* (i.e., $[T]$ contains more than one world), this is no longer the case. For example, suppose that $T = Cn(a \leftrightarrow b, c \leftrightarrow d)$ – where a, b, c, d, are the propositional variables of the language – and let r, r' be the possible worlds $r = Cn(\{\neg a, b, c, d\})$, and $r' = Cn(\{a, \neg b, c, d\})$. It is not hard to see that, although r and r' are different, *Diff*$(T, r) =$ *Diff*$(T, r') = \{a, b\}$. The two worlds r and r', have also another thing in common: they agree on the propositional variables *outside Diff*(T, r). We call such worlds *external T-duals* (for the definition below, recall that P is the set of all propositional variables in the language $\mathcal{L}$):

Definition 3 *Let r, r' be possible worlds, and let T be a theory of $\mathcal{L}$. The worlds r and r' are* external T-duals *iff Diff$(T, r) =$ Diff(T, r') and $r \cap (P -$ Diff$(T, r)) = r' \cap (P -$ Diff$(T, r'))$.*

[9] Notice that in order to increase readability, in this example we are representing worlds as sequences of literals rather than theories; moreover, the negation of a propositional variable p is denoted $\overline{p}$.

Multiple T-duals (external and *internal* ones as we will see later) add more structure to a system of spheres, and render condition (PS) too strong for the general case. The possibility of placing external T-duals in *different* spheres, opens up new ways of ordering worlds that still induce relevance-sensitive revision functions without however submitting entirely to the demands of (PS).

Let us elaborate on this point. Consider a system of spheres S_T centered on the theory T, and let r, r' be any two worlds such that *Diff*$(T, r) \subset$ *Diff*(T, r). Theorem 2 tells us that in the special case of complete theories, to ensure local change (alias, condition (R1)) the world r should be placed (strictly) closer to the center $[T]$ of S_T than r'. In the general case however, and with the aid of external T-duals, one can perhaps afford to be a bit more liberal about the location of r; perhaps all that is needed is that at least one external T-dual r'' of r (and not necessarily r itself) be closer to $[T]$ than r'. It turns out that, in fact, this is pretty much the case, expect that the world r'' "covering" for r (in relation to r') is not just any external T-dual of r but a very specific one: it is the external T-dual of r that agrees with r' on all literals in *Diff*(T, r). We shall call this external T-dual of r, the *r'-cover for r at T*, and we shall denote it by $\vartheta_T(r, r')$.

Definition 4 *Let T be a theory of $\mathcal{L}$, let r, r' be two possible worlds such that Diff$(T, r) \subset$ Diff(T, r'), and let r'' be an external T-dual of r. The world r'' is the r'-cover for r at T iff $r'' \cap$ Diff$(T, r) = r' \cap$ Diff(T, r). We shall denote the r'-cover for r at T by $\vartheta_T(r, r')$.*

A simple example will help to clarify the above definition. Suppose that the language $\mathcal{L}$ is built from the propositional variables a, b, c, d, e, and let T be the theory $T = Cn(a \leftrightarrow b, b \leftrightarrow c, d \leftrightarrow e)$. Let r be the world $r = Cn(a, \neg b, c, d, e)$ and r' the world $r' = Cn(\neg a, b, \neg c, d, \neg e)$. The finest T-splitting is $\{\{a, b, c\}, \{d, e\}\}$. Then according Definition 2, *Diff*$(T, r) = \{a, b, c\}$ and *Diff*$(T, r') = \{a, b, c, d, e\}$. Hence *Diff*$(T, r) \subset$ *Diff*(T, r'). The world r has many external T-duals like $Cn(\neg a, \neg b, c, d, e)$, $Cn(a, b, \neg c, d, e)$, etc. Yet out of all these external T-duals, only one is a r'-cover for r at T, namely the world $\vartheta_T(r, r') = Cn(\neg a, b, \neg c, d, e)$.

As mentioned earlier, the notion of "covering" will be used to weaken condition (PS). In particular, consider the condition (Q1) below:

(Q1) If *Diff*$(T, r) \subset$ *Diff*(T, r') then there is a sphere $V \in S_T$ that contains $\vartheta_T(r, r')$ but not r'.

Condition (Q1) formalizes the intuition mentioned earlier about weakening (PS) with the aid of external T-duals. It is not hard to show that (PS) entails (Q1), and that (Q1) collapses to (PS) when the initial belief set T is complete. Moreover, (Q1) is *strictly* weaker than (PS). To see this, consider the first example in this section; the system of spheres S_T satisfies (Q1) but violates (PS).

Yet despite all its nice properties, condition (Q1) in itself does not suffice to guarantee local change; it seems that from something too strong for (R1) (condition (PS)), we have now moved to something too weak. Consider in particular the following counter-example: the language $\mathcal{L}$ is build over three propositional variables a, b, c, the initial belief set T is $T = Cn(a \leftrightarrow b)$, and the system of sphere S_T centered on $[T]$ is the one represented below:

$$\begin{matrix} abc \\ ab\bar{c} \\ \bar{a}\bar{b}c \\ \bar{a}\bar{b}\bar{c} \end{matrix} \quad \leq \quad a\bar{b}c \quad \leq \quad \bar{a}bc \quad \leq \quad \begin{matrix} a\bar{b}\bar{c} \\ \bar{a}b\bar{c} \end{matrix}$$

In this example all the worlds outside $[T]$ (i.e. in $\mathcal{M}_{\mathcal{L}} - [T]$) differ from $[T]$ on precisely the same propositional variables, namely on $\{a, b\}$. Consequently S_T trivially satisfies (Q1) since its antecedent *Diff*$(T, r) \subset$ *Diff*(T, r') never holds for $r, r' \notin [T]$. Yet despite the compliance with (Q1), the revision function * induced from S_T violates (R1) at T (simply consider the revision of T by $a \wedge \neg b$).[10]

To secure the correspondence with (R1), condition (Q1) needs to be complimented with a second condition, called (Q2). This second condition uses the notion of an *internal T-dual* defined below:

Definition 5 *Let r, r' be possible worlds, and let T be a theory of $\mathcal{L}$. The worlds r and r' are* internal T-duals *iff Diff(T, r) = Diff(T, r'), and $r \cap$Diff(T, r) = $r' \cap$Diff(T, r').*

To give a concrete example of internal T-duals and highlight their difference from external ones, suppose that $\mathcal{L}$ is built over the propositional letters a, b, c, d, and let the initial theory T be $T = Cn(a \leftrightarrow b, c \leftrightarrow d)$. The worlds $r = Cn(\neg a, b, c, d)$ and $r' = Cn(\neg a, b, \neg c, \neg d)$ differ from T on exactly the same propositional variables, i.e., *Diff*(T, r) = *Diff*$(T, r') = \{a, b\}$. Yet r and r' are not external T-duals since outside *Diff*(T, r) the two worlds are not identical. On the other hand, r and r' are identical *inside Diff*(T, r). This makes them internal T-duals.

Clearly, for any theory T and any two worlds r, r', if r and r' are both internal and external T-duals, then they are identical.

We can now proceed with the presentation of condition (Q2), which together with (Q1), brings about the correspondence with (R1). As usual, in the following condition T is an arbitrary consistent theory of $\mathcal{L}$ (possibly incomplete), S_T is a system of spheres centered on $[T]$, and r, r' are possible worlds.

(Q2) If r and r' are internal T-duals, then they belong to the same spheres in S_T; i.e., for any sphere $V \in S_T$, $r \in V$ iff $r' \in V$.

Notice that in the special case that T is complete, no world r has internal or external T-duals (other than itself). Consequently, in that case, (Q1) reduces to (PS), while (Q2) degenerates to a vacuous condition.

The promised correspondence between (R1) and the two conditions (Q1) - (Q2) is given by the theorem below:

Theorem 4 *Let $*$ be a revision function satisfying (K*1) - (K*8), T a consistent theory of $\mathcal{L}$, and S_T a system of spheres centered on $[T]$, that corresponds to $*$ by means of (S*). Then $*$ satisfies (R1) at T iff S_T satisfies (Q1) - (Q2).*

Proof.

$(\Rightarrow)$

Assume that $*$ satisfies (R1) at T. Starting with (Q1), let r, r' be two consistent complete theories such that *Diff*$(T, r) \subset$ *Diff*(T, r'). If *Diff*$(T, r) = \emptyset$, then r is consistent with T, and therefore $r \in [T]$. Then, given that $r' \notin [T]$, (Q1) follows trivially from (S2). Assume therefore that *Diff*$(T, r) \neq \emptyset$. Let $\mathcal{L}'$ be the sublanguage of $\mathcal{L}$ defined over *Diff*(T, r), and let φ be the conjunction of all literals in $\mathcal{L}'$ that hold at r' i.e., $\varphi = l_1 \wedge \cdots \wedge l_k$, where for each $1 \leq i \leq k$, l_i is a literal in $\mathcal{L}'$, and $l_i \in r'$. By (S*), $[T * \varphi] = f_T(\varphi) = [\varphi] \cap C_T(\varphi)$, where $C_T(\varphi)$ is the smallest sphere in S_T intersecting $[\varphi]$. Clearly, since φ is consistent, $f_T(\varphi) \neq \emptyset$. Consider now any world r'' in $f_T(\varphi)$. From (R1) and the construction of φ, it follows that *Diff*$(T, r'') \subseteq$ *Diff*(T, r). Moreover, again by the construction of φ, *Diff*$(T, r) \subseteq$ *Diff*(T, r''). Hence *Diff*(T, r) = *Diff*(T, r''). Moreover, since r'' satisfies φ, $r'' \cap$ *Diff*(T, r) = $r' \cap$ *Diff*(T, r). Finally notice that, because of (R1), $r' \notin C_T(\varphi)$. We have therefore shown that all worlds in $f_T(\varphi)$ are closer to the center of S_T than r', that they all differ from T on exactly the same propositional variables as r, and that within *Diff*(T, r) they agree with r'. What is still left to show in order to prove (Q1) is that there is at least one world in $f_T(\varphi)$ that agrees with r on the propositional variables outside *Diff*(T, r). Let ψ be the conjunction of literals in $\overline{\mathcal{L}'}$ that hold at r. Since r differs from T only in the propositional variables in $\mathcal{L}'$, it follows that $\neg\psi \notin T$. Consequently, by (R1), $\neg\psi \notin T * \varphi$. This again entails that there is at least one world $f_T(\varphi)$, that satisfies ψ and therefore agrees with r on all propositional variables outside *Diff*(T, r) as desired. This concludes the proof of (Q1).

For (Q2), assume on the contrary that it is not true at S_T, and let V be the smallest sphere in S_T that violates (Q2). Then, there exist two worlds r, r' that are internal T-duals such that $r \in V$ and $r' \notin V$. From $r' \notin V$ we firstly derive that *Diff*$(T, r') \neq \emptyset$, and therefore *Diff*$(T, r) \neq \emptyset$, which again entails that $[T] \subset V$. Next, let $\mathcal{L}'$ be the sublanguage of $\mathcal{L}$ defined over *Diff*(T, r). Define φ to be the conjunction of all literals in $\mathcal{L}'$ that hold at r (and therefore also hold at r'). Clearly $C_T(\varphi) \subseteq V$. Moreover, from (R1) and the construction of φ it follows that any world r'' in $f_T(\varphi)$ differs from T on exactly the same propositional variables as r i.e., *Diff*(T, r'') = *Diff*(T, r). In addition, since r'' satisfies φ, r'' and r are in fact internal T-duals, which also makes r'' and r' internal T-duals. Consider now the sentence ψ defined

[10] It is worth noting that in this example, there is only *one* system of spheres S_T whose revision function * satisfies (R1) at T. This is the system containing only the spheres $[T]$ and $\mathcal{M}_{\mathcal{L}}$.

as the conjunction of literals in $\overline{\mathcal{L}'}$ that hold at r'. Since $r' \notin C_T(\varphi)$ (recall that $C_T(\varphi) \subseteq V$), and moreover all worlds in $f_T(\varphi)$ satisfy φ, it follows that no world in $f_T(\varphi)$ satisfies ψ, or equivalently, all worlds in $f_T(\varphi)$ satisfy $\neg\psi$. Hence, $\neg\psi \in T * \varphi$. From (R1) this entails that $\neg\psi \in T$, which again entails that *Diff*$(T, r') \cap \overline{\mathcal{L}'} \neq \emptyset$, leading us to a contradiction. Hence (Q2) is true.

($\Leftarrow$)

Assume that S_T satisfies (Q1) and (Q2). Let T be such that $T = Cn(\chi, \psi)$, for some sentences $\chi, \psi \in \mathcal{L}$ such that $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$. Moreover, let φ be any sentence in $\mathcal{L}_\chi$. If $[T] \cap [\varphi] \neq \emptyset$, then (R1) trivially holds. Assume therefore that $[T] \cap [\varphi] = \emptyset$. Firstly we show that $T \cap \overline{\mathcal{L}_\chi} \subseteq (T * \varphi) \cap \overline{\mathcal{L}_\chi}$. Let γ be any sentence in $T \cap \overline{\mathcal{L}_\chi}$. Assume, contrary to the theorem, that for some r in $f_T(\varphi)$, $\neg\gamma \in r$. Let w be any world in $[T]$. Construct r' as follows: r' agrees with r in $\mathcal{L}_\chi$ and it agrees with w outside $\mathcal{L}_\chi$ i.e., $r' \cap \mathcal{L}_\chi = r \cap \mathcal{L}_\chi$, and $r' \cap \overline{\mathcal{L}_\chi} = w \cap \overline{\mathcal{L}_\chi}$. Clearly, *Diff*$(T, r') \subset$ *Diff*(T, r). Then by (Q1), there is a sphere V smaller than $C_T(\varphi)$ that contains $\vartheta_T(r', r)$ (i.e. the r-cover for r' at T). It is not hard to verify that $\varphi \in \vartheta_T(r', r)$, which however contradicts the fact that $\vartheta_T(r', r) \in V \subset C_T(\varphi)$. Hence $T \cap \overline{\mathcal{L}_\chi} \subseteq (T * \varphi) \cap \overline{\mathcal{L}_\chi}$ as desired.

For the converse, let γ be any sentence in $\overline{\mathcal{L}_\chi}$ such that $\gamma \notin T$. Then there is a world $w \in [T]$ such that $\neg\gamma \in w$. Let r be any world in $f_T(\varphi)$. Construct the world r' as follows: r' agrees with r in $\mathcal{L}_\chi$ and it agrees with w outside $\mathcal{L}_\chi$ (i.e., $r' \cap \mathcal{L}_\chi = r \cap \mathcal{L}_\chi$, and $r' \cap \overline{\mathcal{L}_\chi} = w \cap \overline{\mathcal{L}_\chi}$). Since, as we have shown in the first part of the proof, $T \cap \overline{\mathcal{L}_\chi} \subseteq (T * \varphi) \cap \overline{\mathcal{L}_\chi}$, it follows that *Diff*$(T, r) \subseteq \mathcal{L}_\chi$. Then, by the construction of r', it follows that r and r' are internal T-duals. Consequently, by (Q2), $r' \in C_T(\varphi)$, and since r' satisfies φ, it follows that $r' \in f_T(\varphi)$. Finally notice that, by construction, $\neg\gamma \in r'$. Consequently $\gamma \notin T * \varphi$ as desired. Combining the above we derive that $T \cap \overline{\mathcal{L}_\chi} = (T * \varphi) \cap \overline{\mathcal{L}_\chi}$. ■

Condition (R2) and Systems of Spheres

We now turn to the second side of axiom (P), encoded by condition (R2). As noted previously, a ramification of (R2) is that it introduces dependencies between revision policies associated with *different* theories. Not surprisingly, the condition corresponding to (R2) in the realm of systems of spheres, is one that makes associations between systems of spheres with different centers.

Definition 6 *Let V be a set of worlds in $\mathcal{M}_\mathcal{L}$, and let $\mathcal{L}'$ be a sublanguage of $\mathcal{L}$. By $V/\mathcal{L}'$ we denote the restriction of V to $\mathcal{L}'$; that is, $V/\mathcal{L}' = \{r \cap \mathcal{L}' : r \in V\}$. Moreover, for a system of spheres S_T, by $S_T/\mathcal{L}'$ we denote the restriction of S_T to $\mathcal{L}'$; that is, $S_T/\mathcal{L}' = \{V/\mathcal{L}' : V \in S_T\}$.*

Notice that for any sublanguage $\mathcal{L}'$ of $\mathcal{L}$, $S_T/\mathcal{L}'$ is also a system of spheres. The condition (Q3) below is the semantic counterpart of (R2). As usual, P' is a subset of P, T and T' are theories of $\mathcal{L}$, and S_T, $S_{T'}$ are systems of spheres centered on $[T]$ and $[T']$ respectively:

(Q3) If $\{P', (P - P')\}$ is both a T- splitting and a T'-splitting, and moreover, $T \cap \mathcal{L}^{P'} = T' \cap \mathcal{L}^{P'}$, then $S_T/\mathcal{L}^{P'} = S_{T'}/\mathcal{L}^{P'}$.

The following result shows that (Q3) is the system-of-spheres counterpart of (R2):

Theorem 5 *Let * be a revision function satisfying (K*1) - (K*8), and $\{S_T\}_{T \in \mathcal{K}_\mathcal{L}}$ a family of systems of spheres (one for each theory T in $\mathcal{K}_\mathcal{L}$), corresponding to $*$ by means of (S*). Then $*$ satisfies (R2) iff $\{S_T\}_{T \in \mathcal{K}_\mathcal{L}}$ satisfies (Q3).*

Proof.

($\Rightarrow$)

Assume that $*$ satisfies (R2), and let T, T' be two theories of $\mathcal{L}$. Moreover, assume that for some $P' \subseteq P$, the sets P', $(P - P')$ are a splitting of P relative to both T and T', and $T \cap \mathcal{L}^{P'} = T' \cap \mathcal{L}^{P'}$. If $P' = \emptyset$ or $P' = P$, then (Q3) is trivially true. Assume therefore that $\emptyset \neq P' \subset P$. Then, for some sentences χ, ψ, and ψ', $T = Cn(\chi, \psi)$, $T' = Cn(\chi, \psi')$, $\mathcal{L}_\chi = \mathcal{L}^{P'}$, and $\psi, \psi' \in \mathcal{L}^{(P-P')} = \overline{\mathcal{L}_\chi}$. Consequently, by (R2), for any $\varphi \in \mathcal{L}_\chi$, $(T * \varphi) \cap \mathcal{L}_\chi = (T' * \varphi) \cap \mathcal{L}_\chi = (Cn(\chi) * \varphi) \cap \mathcal{L}_\chi$.

Next consider the systems of spheres S_T and $S_{T'}$ centered on $[T]$ and $[T']$ respectively, and assume that, contrary to Theorem 5, $S_T/\mathcal{L}^{P'} \neq S_{T'}/\mathcal{L}^{P'}$. Without loss of generality, we can assume that $S_T/\mathcal{L}^{P'}$ contains an element that is not in $S_{T'}/\mathcal{L}^{P'}$. Let V be the smallest sphere in S_T such that $V/\mathcal{L}^{P'} \notin S_{T'}/\mathcal{L}^{P'}$. Moreover, let V' be the smallest sphere in $S_{T'}$ such that $V/\mathcal{L}^{P'} \subset V'/\mathcal{L}^{P'}$. Clearly, $[T] \subset V \subset \mathcal{M}_\mathcal{L}$ and $[T'] \subset V' \subset \mathcal{M}_\mathcal{L}$. Consider now a world $r' \in V'$ such that $r' \cap \mathcal{L}^{P'} \notin V/\mathcal{L}^{P'}$. Next, consider a world $r \in V$ such that $r \cap \mathcal{L}^{P'} \notin (\bigcup\{U' \in S_{T'} : U' \subset V'\})/\mathcal{L}^{P'}$. It is not hard to verify that such a world r indeed exists, and moreover, $r \cap \mathcal{L}^{P'} \notin (\bigcup\{U \in S_T : U \subset V\})/\mathcal{L}^{P'}$. Let $l_1, \cdots l_m$ be the literals in $\mathcal{L}^{P'}$ that hold in r, and let $l'_1, \cdots l'_m$ be the literals in $\mathcal{L}^{P'}$ that hold in r'. Finally, let φ be the sentence $\varphi = (l_1 \wedge \cdots \wedge l_m) \vee (l'_1 \wedge \cdots \wedge l'_m)$. Then the smallest sphere intersecting $[\varphi]$ in S_T and in $S_{T'}$, is V and V' respectively. From this we derive that $r'/\mathcal{L}^{P'} \notin f_T(\varphi)/\mathcal{L}^{P'}$ and $r'/\mathcal{L}^{P'} \in f_{T'}(\varphi)/\mathcal{L}^{P'}$. Consequently, $(T * \varphi) \cap \mathcal{L}_\chi \neq (T' * \varphi) \cap \mathcal{L}_\chi$, which however contradicts (R2).

($\Leftarrow$)

Assume that $\{S_T\}_{T \in \mathcal{K}_\mathcal{L}}$ satisfies (Q3) and let T be a theory of $\mathcal{L}$ such that for some $\chi, \psi \in \mathcal{L}$, $T = Cn(\chi, \psi)$ and $\mathcal{L}_\chi \cap \mathcal{L}_\psi = \emptyset$. Let T' be the theory $T' = Cn(\chi)$. Clearly, $T \cap \mathcal{L}_\chi = T' \cap \mathcal{L}_\chi$, and therefore by (Q3), $S_T/\mathcal{L}_\chi = S_{T'}/\mathcal{L}_\chi$. Next, consider any sentence $\varphi \in \mathcal{L}_\chi$, and let

$C_T(\varphi)$ and $C_{T'}(\varphi)$ be the smallest spheres in S_T and $S_{T'}$ respectively intersecting $[\varphi]$. We show that $C_T(\varphi)/\mathcal{L}_\chi = C_{T'}(\varphi)/\mathcal{L}_\chi$. Assume on the contrary that $C_T(\varphi)/\mathcal{L}_\chi \neq C_{T'}(\varphi)/\mathcal{L}_\chi$. Since $S_T/\mathcal{L}_\chi = S_{T'}/\mathcal{L}_\chi$, without loss of generality we can then assume that $C_T(\varphi)/\mathcal{L}_\chi \subset C_{T'}(\varphi)/\mathcal{L}_\chi$. Moreover, again from $S_T/\mathcal{L}_\chi = S_{T'}/\mathcal{L}_\chi$, we have that $C_T(\varphi)/\mathcal{L}_\chi \in S_{T'}/\mathcal{L}_\chi$. Let V be the sphere in $S_{T'}$ whose restriction to $\mathcal{L}_\chi$ is equal to $C_T(\varphi)$ i.e., $V/\mathcal{L}_\chi = C_T(\varphi)/\mathcal{L}_\chi$. Clearly, $V \subset C_{T'}(\varphi)$. On the other hand however, since $\varphi \in \mathcal{L}_\chi$ and $C_T(\varphi)$ contains as least one φ-world, it follows that V also contains a φ-world. This of course contradicts the fact that $C_{T'}(\varphi)$ is the smallest sphere in $S_{T'}$ intersecting $[\varphi]$, and proves that $C_T(\varphi)/\mathcal{L}_\chi = C_{T'}(\varphi)/\mathcal{L}_\chi$. Consequently, $(T * \varphi) \cap \mathcal{L}_\chi = (T' * \varphi) \cap \mathcal{L}_\chi$ as desired. ■

Putting together the results reported in Theorems 1, 4, and 5, we obtain immediately the following theorem that provides possible-world semantics for (the strong version of) axiom (P):

Theorem 6 *Let $*$ be a revision function satisfying (K*1) - (K*8) and $\{S_T\}_{T \in \mathcal{K}_\mathcal{L}}$ a family of systems of spheres (one for each theory T in $\mathcal{K}_\mathcal{L}$), corresponding to $*$ by means of (S*). Then $*$ satisfies (P) iff $\{S_T\}_{T \in \mathcal{K}_\mathcal{L}}$ satisfies (Q1) - (Q3).*

Conclusion

The main contribution of this paper is Theorem 6 that provides system-of-spheres semantics for Parikh's axiom (P). What is quite appealing about this result is that the semantic conditions (Q1) - (Q3) that characterize axiom (P) are quite natural constraints on similarity between possible worlds. In fact, conditions (Q1) - (Q2) essentially generalize a measure of similarity that predates axiom (P), and was motivated independently in the context of Reasoning about Action by Winslett. This intuitive nature of the semantics is more evident in the special case of consistent complete belief sets. An interesting by-product of our study is the identification of the two possible readings of axiom (P), both of which are plausible depending on the context.

It should be noted that apart from Winslett, other authors have also made specific proposals for measuring distance between possible worlds (see for example, (Borgida 1985), (Dalal 1988), and (Satoh 1988)).[11] It would be a worthwhile exercise to investigate whether any of these measures of distance also yield some kind of "local change effect" for their associated revision functions.

Acknowledgements

We would like to thank the anonymous referees for many useful comments on this article. The first author is also grateful to the School of Computer Science and Engineering at the University of New South Wales for funding his visits to Sydney and thus enabling this joint work.

[11]For a more general study on the notion of distance in belief revision, refer to (Lehmann, Magidor, & Schlechta 2001).

References

AIJ. 1997. Special issue on relevance. *Artificial Intelligence* 97.

Alchourron, C.; Gardernfors, P.; and Makinson, D. 1985. On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic* 50:510–530.

Borgida, A. 1985. Language features and flexible handling of exceptions in information systems. *ACM Transactions on Database Systems* 10:563–603.

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

Chopra, S., and Parikh, R. 2000. Relevance sensitive belief structures. *Annals of Mathematics and Artificial Intelligence* 28(1-4):259–285.

Chopra, S.; Georgatos, K.; and Parikh, R. 2001. Relevance sensitive non-monotonic inference for belief sequences. *Journal of Applied Non-Classical Logics*.

Dalal, M. 1988. Investigations into a theory of knowledge base revisions: Preliminary report. In *Proceedings of the 7th AAAI Conference*, 475–479.

Gardenfors, P., and Makinson, D. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of Theoretical Aspects of Reasoning about Knowledge*, 83–95. Morgan-Kaufmann.

Grove, A. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17:157–170.

Lehmann, D.; Magidor, M.; and Schlechta, K. 2001. Distance semantics for belief revision. *Journal of Symbolic Logic* 66(1):295–317.

Parikh, R. 1999. Beliefs, belief revision, and splitting languages. In Lawrence Moss, J. G., and de Rijke, M., eds., *Logic, Language, and Computation Volume 2 - CSLI Lecture Notes No. 96*. CSLI Publications. 266–268.

Peppas, P., and Wobcke, W. 1992. On the use of epistemic entrenchment in reasoning about action. In *Proceedings of the 10th European Conference on Artificial Intelligence*, 403–407. John Wiles and Sons.

Peppas, P.; Foo, N.; and Nayak, A. 2000. Measuring similarity in belief revision. *Journal of Logic and Computation* 10(4).

Peppas, P. 1994. *Belief Revision and Reasoning about Action*. Ph.D. Dissertation, Basser Department of Computer Science, University of Sydney.

Satoh, K. 1988. Nonmonotonic reasoning by minimal belief revision. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 455–462.

Winslett, M. 1988. Reasoning about action using a possible models approach. In *Proceedings of the 7th AAAI Conference*, 89–93.