A Revised Version:
Belief Revision and Epistemic Acts

MSc Thesis (*Afstudeerscriptie*)

written by

Jonathan A. Zvesper
(born November 4th, 1981 in Norwich, England)

under the supervision of **Prof Krister Segerberg** and **Dr Eric Pacuit**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
January 18th, 2007	Prof Johan van Benthem
	Dr Eric Pacuit
	Prof Krister Segerberg
	Prof Frank Veltman

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Acknowledgements

It is customary to open this sort of dissertation with some words of thanks to those who have helped one along the path to achieving it. Even if it were not, I like to think that I would find the space in these mostly rather dry pages to do so. For in this case, things would not have been possible without the help of a few, and would not have been so enjoyable without the presence of many.

More specifically, to both of my supervisors, Krister and Eric, I owe thanks. Krister has revealed a great amount of knowledge, explaining (often patiently) many of the subtleties involved in the topics addressed in this thesis. Beyond that much more: He and his wife Anita have proved themselves very successful hosts and good company at dinners and other events. I hope to become as sharp, insightful and open-minded as Krister is by the time I am seventy. Sometimes he even laughed at my jokes. Eric has suggested a number of directions for the research contained (or not contained) in this thesis, and has provided a great deal of enthusiasm. Even better, he always laughed at my jokes.

A number of fruitful discussions with Johan, in the period before I started working with Krister and Eric, should be acknowledged for having been particularly inspiring. A number of other people around the Institute for Logic, Language and Computation deserve mentions, for being friends or colleagues or both. They will not all get them — I trust that they know who they are — but I would like in particular to thank the dedicated and efficient administrative staff: Tanja, Marjan, Jessica, Ingrid, even René. Their contribution may seem very indirect, but after two years in a French academic environment I have become very sensitive to bureaucratic nonsense, and am pleased that they actually go out of their way to make all aspects of student life at the ILLC as easy and pleasant as they can. Although officially for Benedikt being a member of the administrative staff is only a small part of his employment, it is in this capacity that I have encountered him the most, and I would also like to thank him with the same sentiments for his efficiency and the sheer amount of time puts in to making things run smoothly at the ILLC.

I also still want to thank Olivia for being her, and for coming to Amsterdam. Though the emphasis is on the former, because perhaps we would still have met elsewhere.

Contents

1 Introduction

This thesis is about mathematical models for representing changes in beliefs of groups of agents.

We do not offer directly an analysis of what a belief is. Instead we take as a natural starting point that a belief is something which an agent accepts as true. A richer view of belief than this 'binary' one will in effect be presented, but to repeat: This can only indirectly be read as an analysis of what a belief is.

One important theme of this thesis is the status of the triggers of belief changes, as what have been called 'epistemic actions'; for example actions of observation or of communication.

There is no principal question or thesis to this thesis. Different ways to represent the richness of belief states in a multi-agent setting are explored, and the theme of epistemic actions recurs. A number of original technical results are presented[1]. Each of §3 – 7 offer, in lieu of a conclusion[2], though not necessarily exclusively at the end of each chapter, some precisely demarcated directions for future research.

In §2, we start by introducing the domain of belief revision in the single-agent case, and where the information involved in the revision does not concern the agent's epistemic states ([1], [16]). We explain the inadequacy of this approach for the multi-agent and introspective cases. We then summarise some logics for reasoning about *knowledge* changes in groups of agents ([14], [4]) that overcome those inadequacies in the case of knowledge change. At the end of the chapter we introduce a basic semantic framework (of orderings over interpretations) that enriches the binary view of belief states, and which will recur in future sections of the thesis.

In §3, we consider one topic of multi-agent belief revision, viz. the merging of belief states when these are considered purely from the semantic perspective ([22]). The topic for further research is to give a full syntactic account of such mergings. We will propose some initial thoughts, and some quite concrete things, in that section.

[1] I wavered between relegating proofs to appendices, and chose to include them in the main body of the text. I hope that the readability does not suffer too much for this decision.

[2] This thesis avoids the "La bêtise" which Flaubert once said "consiste à vouloir conclure"; Nonetheless it also avoids following too literally the aesthetic dictate of his remark, again in a letter to Louise Colet, that "Les oeuvres les plus belles sont celles où il y a le moins de matiére". It is in any case open to the criticism of resembling the young M. Bovary's hat.

Then in §4, we consider some logical languages (from [31], [7]) that bring together the various parts of the first chapter. That is, they express beliefs and changes in beliefs amongst groups of agents. We provide a completeness proof, in §5, of a new logic introduced in that chapter.

In §6 we discuss an algebraic approach to multi-agent belief revision ([5]). Those algebraic structures enable us (the modelers) to model agents revising their theories with respect to the epistemic actions themselves, and not just with respect to the way the world is.

When reasoning about epistemic actions in the context of beliefs rather than knowledge, even more pragmatic reasoning about those actions might take place, and this is a useful way to start modeling that reasoning. We discuss this issue in §7, where we consider how to model those additional aspects to the reasoning that agents might make about actions. That last chapter is the most diaphanous: It is rather speculative, with no concrete results, and is included mainly as a pointer to other things which are possible in the context of reasoning about belief change.

2 Beliefs and Knowledge-Changing Actions

2.1 Introduction

Given a formal language L that is intended to describe a certain environment, and some syntactic rules of inference for L, the beliefs B of a rational agent about that environment can be represented by a set Γ_B of L-sentences that is closed under those rules. Then we say that the sentences that the agent thinks are true are those that are in Γ_B. Or, equivalently, those beliefs can be represented by the class of interpretations $\mathcal{K}_B$ (possible configurations of the environment) that make exactly the members of Γ_B true. In this case we say that the agent thinks that the actual configuration of the environment is a member of $\mathcal{K}_B$. Belief change is then represented by considering a new set of sentences Γ_C (class of interpretations $\mathcal{K}_C$).

In the simplest case, to represent the agent's learning of a fact F that she did not previously disbelieve, one way to reason about what occurs is simply to add to Γ_B the formulae corresponding to F and close under inference (semantically: to remove from consideration those members of $\mathcal{K}_B$ in which F does not hold). Complications arise when F is not consistent with the agent's beliefs; i.e. when she did previously disbelieve F. We will call the former kind of situation belief *expansion*; the latter kind of situation is referred to as belief *revision*.

There are also complications that arise when we consider cases where the environment includes the beliefs of a group of such rational agents, because the environment changes under the influence of such belief change. To model such situations we should think of the belief change, or of whatever brings it about, as being an action upon the environment. We will see in section 2.2 why the syntactic postulates proposed by Alchourrón Gärdenfors and Makinson can legitimately be thought inappropriate in these cases. In §2.3.2 and 2.3.3, we will look at an approach to manage these complications in the case where what is at stake is changes in *knowledge*, which is (at least) veridical belief.

2.2 Syntactic Approach

2.2.1 AGM Postulates

Any discussion of belief revision must take into account the work of Alchourrón, Gärdenfors and Makinson [1]. Their work involved a formalisation of the notion of a rational way to change a theory Γ when one is forced to accept some piece of information q which might not be consistent with Γ. Evidently the rational thing to do is to retain *as much as one can* of Γ without being inconsistent with q. [1] presents formal investigations into what can be meant by that "*as much as one can*".

We take a theory to be a set of sentences closed under the logical consequence relation $\vdash$, and denote by $\Sigma_{\mathcal{L}}$ the set of theories over the language $\mathcal{L}$.[1] Now we can express, using $(\Gamma)^{\vdash}$ to denote the closure by $\vdash$ of Γ and $\perp$ to represent *falsum*, the axiomatic postulates which [1] proposed. Given a theory Γ, we say that $\dotplus : \mathcal{L} \to \Sigma_{\mathcal{L}}$ is an AGM revision operator for Γ when the following conditions hold for any formula ϕ (we write $\Gamma \dotplus \phi$ instead of $\dotplus(\phi)$):

($\dotplus$1) $\Gamma \dotplus \phi \vdash \phi$

($\dotplus$2) $\Gamma \nvdash \neg\phi \Rightarrow \Gamma \dotplus \phi = (\Gamma \cup \phi)^{\vdash}$

($\dotplus$3) $\nvdash \neg\phi \Rightarrow \Gamma \dotplus \phi \nvdash \perp$

($\dotplus$4) $\vdash \phi \equiv \psi \Rightarrow \Gamma \dotplus \phi = \Gamma \dotplus \psi$

($\dotplus$5) $\Gamma \dotplus (\phi \wedge \psi) \subseteq ((\Gamma \dotplus \phi) \cup \{\psi\})^{\vdash}$

($\dotplus$6) $(\Gamma \dotplus \phi) \nvdash \neg\psi \Rightarrow ((\Gamma \dotplus \phi) \cup \{\psi\})^{\vdash} \subseteq \Gamma \dotplus (\phi \wedge \psi)$

2.2.2 Explanation of the Postulates

We briefly explain these conditions, to see their intuitive appeal in the kind of theory change under consideration in [1]; we will then see reasons for rejecting some of the postulates in a multi-agent context, where the revision operator becomes something different.

The first of these conditions says that new information is always believed after the operation of revising by it. The aim of a theory of belief revision is to give an account of the way beliefs, or a theory, change when one actually accepts some piece of incoming information. Thus this is a seemingly unquestionable postulate.

[1] The definition of a theory as such a set of sentences might seem hasty, and indeed in defining the type of $\dotplus$ we have already subsumed the very first postulate of [1], which says that $\Gamma \dotplus \phi$ is a theory. We return to the nature of a theory (or rather, a belief state) later.

In the special case of belief expansion, that is when the new information ϕ is not inconsistent with one's actual beliefs Γ, $(\dot{+}2)$ enforces that one should keep all of one's actual beliefs and integrate only the new information (and its consequences) with them.

The third condition says that one can always integrate new information unless it is inconsistent by itself. $(\dot{+}4)$, which like $(\dot{+}1)$ would retain precisely the same force, in the presence of $(\dot{+}2)$, if it were made conditional on the inconsistency of $\Gamma \cup \{\phi\}$, says that logically equivalent incoming information is to be treated equivalently. $(\dot{+}5)$ says that if one revises with ϕ and then adds ψ to one's beliefs, then one can infer at least as much as if one had in the first place revised by $\phi \wedge \psi$.

The last condition has the effect that in case one gets a consistent set of beliefs by first revising by ϕ and then adding ψ to one's beliefs, then performing those two steps has exactly the same effect as revising by $\phi \wedge \psi$.

2.2.3 Inappropriateness of the Axioms

It is not surprising that these axioms do not hold when we consider a completely different operation, the kind of multi-agent revision that we envisage, in which a revision is to be thought of as an epistemic action, perhaps carried out by (a number of) those agents. However, we nonetheless explain with some illustration which of the postulates (all but $(\dot{+}3)$ and $(\dot{+}4)$) fail, and why. We relate each of the reasons to the fact that in our setting, revision is to be modeled as an action that is performed.

Closed Systems

The first reason for rejecting the postulates as applying to actions in a multi-agent setting we will only outline here; it will be taken up again in chapter 7. It is that the multi-agent setting gives us the facility to think of the revision process, the events and their associated belief dynamics, as a closed system in a way which does not make sense from the AGM perspective. $(\dot{+}1)$, and to a lesser extent $(\dot{+}2)$ and $(\dot{+}6)$, are suitable for characterising a situation in which some outside force which must be obeyed, and is able to make itself perfectly understood and recognised as the force which must be obeyed, orders the agent(s) to see to it that they believe a certain proposition about their environment. In situations where one is told something, one might not believe it, or one might believe part of it. (A vivid example of this is offered by Fermé and Hansson [13], who describe a child telling a parent who arrives home that a dinosaur just broke the vase in the living-room. The parent will in usual cases screen the information, revising her belief that the vase in the living-room is intact, but not going so far as to revise her belief that there are no vase-breaking dinosaurs in the vicinity. We will, at least from §2.3 onwards, take the objects of revision to be *propositions*, i.e. semantic objects. This sort of dinosaur-vase topic would require a syntactic analysis, and therefore is not something that we find the space to explore further here.)

So the first reason, which we call 'thinness of actions', is simply that we might want to do more with our notion of revision than the AGM postulates allow, considering revisions as actions of communication between the agents, in a closed system. Revision as given in the postulates is not an action recognised by the implicit agent on which it operates (the theory Γ). In §7.3, we will consider a system in which the first postulate is not rejected on the basis of this reason, by formulating such message-passing actions as more complex formulae. But for now, we turn to more concrete objections to applying the axioms in such a multi-agent context.

Introspection and Empathy

The kind of objection to be raised now involves an important idea, one that that will recur in later sections. There are two parts to it. The first comes from the fact that the environment being modeled should, on pain of triviality given the situations we envisage modeling, contain (at least some of) the beliefs of the agents present in that environment. The second is that when revision can be about those doxastic parts of the environment, then there is of necessity some temporal aspect to the environment. We will now make both of these two problems clearer.

If the agents do not reason about each others' beliefs (and perhaps their own), then, aside from the brief discussions of chapter 3, there seems little reason to say that the agents are in the same environment, as we cannot represent very much information about the situation[2]. So we want our formal language $\mathcal{L}$ to allow us to attribute meta-beliefs to the agents. What exactly do we mean by this? If a formula ϕ is a member of a set Γ_j representing j's beliefs, then we say that j believes ϕ. That is, the formulae of L come to express j's beliefs when they are members of Γ_j. Now suppose that the formula ϕ already expressed a belief: say it was of the form $B_j\psi$, where B_j is, in the tradition of Hintikka [17], a modal operator true precisely of those formulae that j believes. Then when ϕ is a member of Γ_j, it expresses a metabelief: in this case, j's belief that j believes that ψ.

The first kind of objection is to the postulates ($\dotplus 2$) and ($\dotplus 5$). We will discuss the objection, which we call 'retention of disbelief', in order to refer to it later, with respect to ($\dotplus 2$), and then simply explain how it occurs with respect to ($\dotplus 5$).

Suppose that neither p nor $\neg p$ are in Γ_j, and $\neg B_j p \in \Gamma_j$. Now since Γ is closed under logical consequence, both p and $\neg p$ are consistent with Γ_j. So by ($\dotplus 2$), $\Gamma \dotplus p = (\Gamma \cup \{p\})^{\vdash}$. That means that $\neg B_j p \in \Gamma$, so after j has revised with p, j has come to believe that p, but not that she believes that p. Even if the agent happened to be positively introspective before the revision (i.e. $p \in \Gamma_j \Rightarrow B_j p \in \Gamma_j$), she is no longer so afterwards. This is perhaps not a problem, because we might not be concerned with agents' beliefs about

[2] Cf. The discussion of an epistemic representation thesis in section 2.3.2 below.

their own beliefs, but now suppose that we were considering Γ_i and not Γ_j. Then clearly we cannot represent the epistemic action of mutual discovery of the truth of the proposition p by i and j as a revision by p.

Again, this point is not surprising, but it is worth making it clear. A revision by a proposition ϕ must change more than just the 'value' of ϕ in an agent's belief base, but surely should be an action with further effects, even in the case of a single agent. (Notice also that a similar problem occurs with ($\dotplus$5): Suppose that Γ_i is consistent; then by ($\dotplus$2), $\Gamma_i \dotplus \top = \Gamma_i$. Then an instantiation of ($\dotplus$5) is $\Gamma_i \dotplus q \subseteq (\Gamma_i \cup \{q\})^{\vdash}$. Now suppose that $\Gamma_i \not\vdash B_j q$; then unless $\Gamma_i \vdash q \supset B_j q$ then $\Gamma_i \dotplus q \not\vdash B_j q$.)

Moore Problems

The other objection ('Moore-ish objection'), which complains about ($\dotplus$2) and ($\dotplus$6), is related to the so-called 'Moore Sentences', named after G. E. Moore, who observed the absurd nature of an utterance of the form

(M) *I don't believe it but it's raining outside.*

Consider a sentence of the following form:

(M') *You don't believe it, but it's raining outside.*

We would model such a situation by saying that the agent j is made to accept the non-absurd $(r \wedge \neg B_j r)$. The absurdity comes not from that formula, but from the attribution of the formula to j's doxastic state. Yet an instantiation of ($\dotplus$1) is $F \dotplus (r \wedge \neg B_j r) \vdash r \wedge \neg B_j r$. Thus after being forced to accept the (non-absurd) (M'), the agent whose beliefs are being modeled would have (M) as a belief.

This problem is intimately related to the idea that communications and other epistemic activities are actions which change the environment. Assertions are usually said to have a semantic content, or literal meaning, and a pragmatic force or effect. The literal meaning of an utterance is often said (at least since Tarski) to be given by 'truth-conditions', whereas the pragmatic force, which is not part of the literal meaning, is derived from the fact that the utterance is an action, in some context and presumably with some goals.

2.2.4 Iteration of Revision

The *AGM* postulates say nothing explicitly about iterations of revision: There are no postulates of the form $(\Gamma \dotplus \phi) \dotplus \psi$. In the next chapter, specifically in §3.3.2, we will see a natural way to iterate revisions, and we will also look more closely at a number of different methods for iterated revision in §4.2.2.

2.3 Semantic Approach

In §2.1, we mentioned two equivalent ways to represent the epistemic states of agents. The first was syntactically, as sets of sentences, and this is the approach we have seen so far. We now look at two ways to represent beliefs of agents semantically, with respect to interpretations of those sentences.

The first (§2.3.1–2.3.3) is properly multi-agent: the beliefs of the agents are part of the environment. Those representations, based on relational semantics and a monadic modal operator, following the work of Hintikka [17], are very well-suited for reasoning about *knowledge* and its dynamics, but not so well-suited for reasoning about beliefs and theirs.

The distinction between knowledge and belief that is important here is that between 'hard' and 'soft' information (cf. [7]). Hard information leads to knowledge, is irrevocable. When you move your pawn to $Q4$, unless I am the most hardened of skeptics, someone modeling my epistemic state would be justified in modeling part of the information conveyed by your action as a hard information change: I know that you have moved your pawn to $Q4$. However, there is also softer information conveyed by that action: I might read something about your game plan, which I could later readily abandon. When we say that relational semantics are not well-suited for modeling belief change, it is because there is no facility to model such soft, abandonable information changes. Yet those information changes are clearly of interest, even if they are primarily pragmatic. Pragmatic in two senses: Firstly, in the sense that they can, like Grice's pragmatic implicatures, be abandoned without invalidating the framework in which they take place. And secondly in the sense that sometimes one must act in situations where one simply lacks enough hard information to make a decision about what is the best action. In these situations soft information is needed. Knowledge is idealistic, with perfectionist standards, but beliefs are more pragmatic: Just as the moral nihilist might say that there's no ultimate point to anything (there's no hard information), nonetheless she will in her life take certain decisions (she will exploit soft information).

The second semantical representation (§2.3.4) involves only a single (implicit) agent, but is closely connected to the AGM postulates: a single agent's belief state is represented not just as a single *set* of interpretations/sentences, but as what we will call a belief *state*, in that it carries the information about how the agent would change her beliefs if forced to revise them. In later sections, we will consider a number of richer modal languages that are needed to reason about this second kind of semantics in ways similar to those used in §2.3.1 – 2.3.3.

2.3.1 Relational Semantics

We present a standard semantics for n-agent epistemic logic over the propositional language $\mathcal{L}$ over the alphabet Φ of propositional variables. We call this 'relational' even though for expository reasons we lift the usual relations to their corresponding functions. Assuming W to be an arbitrary non-empty set, whose elements we will refer to as "points", we use the following definitions:

Definition 2.3.1 (Relational Local View Function):
A *relational local view function* over W is a function $f : W \to 2^W$.

A local view function is the way in which agents are characterised in relational semantics: Each agent is associated with a relational local view function that is used to give the agent's epistemic state. The intuition behind such a function is that, assuming some point u to be the point that holds, then $f_a(u)$ returns the points which the agent a considers possible. There are many questions that could be raised concerning what is meant here by considering something possible, and we would get a different notion of an epistemic state depending on what notion was intended. The best way to think of the members of $f_a(w)$ are as those points which, at the point w, agent a has not ruled out as candidates for being the point at which the agent actually is.

Definition 2.3.2 (n-agent Relational Frame):
An *n-agent relational frame* over W is an $n+1$-tuple $\langle W, f_0 \dots f_{n-1} \rangle$, where for each $i \in n$, f_i is a relational local view function over W.

Definition 2.3.3 (n-agent Relational Model):
An *n-agent relational model* over W is an n-agent relational frame over W furnished with a *valuation function* $V : W \to 2^\Phi$.

We write $w \models_{\mathfrak{M}} \phi$ to mean that ϕ is true at the point w in the model $\mathfrak{M}$; where the model intended is evident from the context, we just write $w \models \phi$. And we write $[\![\phi]\!]$ to mean the set of points w such that $w \models \phi$. There follows the standard recursive satisfaction definition for truth at a point:

1. $w \models p \Leftrightarrow p \in V(w)$ for $p \in \Phi$
2. $w \models \neg\phi \Leftrightarrow w \not\models \phi$
3. $w \models \phi \wedge \psi \Leftrightarrow (w \models \phi \;\&\; w \models \psi)$
4. $w \models B_i\phi \Leftrightarrow f_i(w) \subseteq [\![\phi]\!]$

We call an n-agent model in which one point is singled out a 'pointed n-agent model'. How do such models relate to the sets-of-sentences view of theories?

Definition 2.3.4 (Global Theory of a Point):
The *global theory* of w, $\Gamma^w = \{\phi | w \models \phi\}$

The global theory of w is just those sentences which are true at w, and is itself a logically closed set of sentences; indeed, a maximally consistent such set: for any $\phi \in \mathcal{L}, \phi \in \Gamma^w \Leftrightarrow \neg\phi \notin \Gamma^w$. That is supposed to be the state of the environment at w. But for each agent, at each point w, we can also define the way that agent sees the environment:

Definition 2.3.5 (Local Theory of a Point):
i's *local theory* of w, $\Gamma_i^w = \{\phi | w \models B_i\phi\}$

Remark 2.3.1:
$\Gamma_i^w = \bigcap_{w' \in f_i(w)} \Gamma^{w'}$

An agent's local theory, then, is again a logically closed set of sentences, though not in general maximal as the agent will usually have some uncertainties, nor in the general case consistent as we have not ruled out that $f_i(w) = \varnothing$. The agent's local theory at a point is of the same kind, then, as we assumed theories to be when introducing the *AGM* postulates.

An n-agent model is intended to be a compact representation of a certain class of 'social epistemic' situations. Baltag and Moss have suggested that the appeal of epistemic logic is that this representation is such that "all intuitive judgements concerning [certain epistemic aspects of a given social situation] correspond to formal assertions concerning [a given pointed n-agent model], and vice-versa.". We give the full quotation of what we will call the 'epistemic state representation thesis':

Thesis 1 (Epistemic State Representation):
"Let s be a social situation involving the intuitive concepts of knowledge, justifiable beliefs and common knowledge among a group of agents. Assume that s is presented in such a way that all the relevant features of s pertaining to knowledge, beliefs and common knowledge are completely determined. Then we may associate to s a mathematical model S. (S is an [n-agent relational model].) The point of the association is that all intuitive judgements concerning s correspond to formal assertions concerning S, and vice-versa." ([3], p.166)

In a relational model, each point is supposed to represent a given way the environment could be. It is important to note that points are not in general individuated by the propositional variables that they make true; indeed, even the semantics given is more fine-grained, making a point's relation to other points, given by the f_i's, of significance. This is because the 'environment' includes the epistemic attitudes of the agents. Such points could more properly be called 'possible worlds' were it the case that they *included* all of the information about that configuration. Most[3] relational models $\langle W, f_0 \dots f_m, V\rangle$ are seemingly obviously mathematically equivalent to a collection $\{s_w\}_{w\in W}$, where $s_w = \langle V(s), s_{f_0(w)} \dots s_{f_m(w)}\rangle$. Yet these latter constructions are mathematically, and perhaps conceptually, unhappy creatures, as they can, and often will be, non-well-founded. To take the smallest example: the relational model $\langle\{*\}, \{\langle *, \{*\}\rangle\}, \{\langle *, \emptyset\rangle\}\rangle$ is equivalent, via this translation, to $\{x\}$, where $x = \langle\emptyset, x\rangle$. Thus for philosophical, and perhaps mathematical, reasons, we present things differently, giving each point a name and then stripping it of its non-well-foundedness by lifting what we might call the 'ground state' of the world to the image by the valuation function, and the 'belief states' of the world to the f_i's, the 'local view functions'.

Another part of the epistemic representation thesis is that epistemic actions such as announcements and discoveries have mathematical correlates in just the same way. We will see the more complex structures to which Baltag and Moss apply that thesis in section 2.3.3. For now, we look at a logic that gives some account of simple actions – public announcements or discoveries.

2.3.2 Announcements – Point Elimination

In AGM theory, revision, which is implicitly by a single agent, occurs through the forced acceptance of a single sentence (formula). This is a kind of hard information change, and following [14], we can model such changes, in the case of knowledge-oriented relational semantics, in an elegant and straightforward way. When an announcement of a true statement is made, and known by each agent to be made truthfully, then each agent can eliminate the points where the statement could not have been truthfully made from the range of all of their local view functions. Furthermore, we as modelers need not add points in order to represent faithfully the situation – to abide by a slightly extended version of the epistemic representation thesis. This is because there are no new possibilities any agent considers, for example representing the fact that she is unsure whether another agent also leaned the truth of the statement (and the fact of the action's taking place).

Thus a simple action of public announcement of a formula ϕ can be represented as an operation on models $-[!\phi]$, which maps an n-agent relational model $\mathfrak{M}$ to its restriction to $[\![\phi]\!]_{\mathfrak{M}}$, i.e. the submodel generated by the points at which, in the original model, ϕ was true. The complete logic of such models, including a modality for announcements $[\phi]$ in the language, is given in [14]. In fact, the logic given in [14] provides for actions of the form $[\mathcal{G}!\phi]$, representing an announcement of ϕ to the group of agents $\mathcal{G}$. We call this logic Public Announcement Logic (PAL). Its axiomatisation is a reduction to the static

[3]We say 'most' here because most accounts of non-well-founded set theory identify bisimilar sets, so that for example two disjoint single reflexive points will become, via such a construction, a singleton. It should be clear though that every model without any two points that have the same theories (i.e. that make the same modal formulae true) will yield an isomorphic non-well-founded set model.

epistemic logic. So assuming axioms for the belief modality, the following 'reduction axioms' are given in [14]:

(GG1) $[\mathcal{G}!\phi]p \equiv (\phi \supset p)$ for any propositional variable p;

(GG2) $[\mathcal{G}!\phi]\neg\psi \equiv \neg[\mathcal{G}!\phi]\psi$;

(GG3) $[\mathcal{G}!\phi](\psi \wedge \chi) \equiv ([\mathcal{G}!\phi]\psi \wedge [\mathcal{G}!\phi]\chi)$;

(GG4) $[\mathcal{G}!\phi]B_i\psi \equiv (\phi \supset B_i(\phi \supset [\mathcal{G}!\phi]\psi))$ for $i \in \mathcal{G}$;

(GG5) $[\mathcal{G}!\phi]B_i\psi \equiv B_i\psi$ for $i \notin \mathcal{G}$.

It is clearly (GG4) that is the interesting axiom. A strength of this logic is that it captures correctly our intuitions about how beliefs change under epistemic actions. In particular, the Moore problems of §2.2.3 are well-handled. For example, (GG4) shows that after an announcement that $p \wedge \neg B_a p$, a does believe that p:

$$\begin{aligned}
[\{a\}!(p \wedge \neg B_a p)]B_a p \quad & \equiv p \wedge \neg B_a p \wedge B_a((p \wedge \neg B_a p) \supset [\{a\}!(p \wedge \neg B_a p)]p) \quad \text{(GG4)} \\
& \equiv p \wedge \neg B_a p B_a((p \wedge \neg B_a p) \supset p) \\
& \text{(by (GG1) and normality of the static logic)} \\
& \equiv p \wedge \neg B_a p
\end{aligned}$$

2.3.3 Action Models

In order to model more complex actions than simply announcements, actions themselves can be considered to have similar a epistemic structure to them as the relational models (which we will now call 'state models') above. These developments were made by [14] and [4]. Consider a non-empty set A, whose members we will call 'atomic actions'.

Definition 2.3.6 (Precondition Function):
$\mu : A \to \mathcal{L}$ is a *precondition function*.

A precondition function specifies a formula that *must* be true in order for an action to take place.

Definition 2.3.7 (n-agent Action Frame):
An *n-agent action frame* is an $n+1$-tuple $\langle A, f_0 \dots f_{n-1}\rangle$, where for each $i \in n$, f_i is a relational local view function over A. We refer to the elements of A not as 'points' but as 'atomic actions'.

Definition 2.3.8 (n-agent Action Model):
An *n-agent action model* is an n-agent action frame furnished with a precondition function.

We then define 'product' update: Given an n-agent state model $\mathcal{S}$ and an n-agent action model $\mathcal{A}$, it returns $\mathcal{S} \otimes \mathcal{A}$, which is either the empty set or an n-agent state model.

Definition 2.3.9 (Product Update):
$\langle W, \{f_i\}_{i \in n}, V \rangle \otimes \langle A, \{g_i\}_{i \in n}, \mu \rangle =$
$\langle W \otimes A = \{\langle w, \delta \rangle \in W \times A | w \models \mu(\delta)\}, \{fg_i\}_{i \in n}, \lambda.\langle w, \delta \rangle V(w) \rangle,$
where for each $i \in n$, $\begin{array}{rccl} fg_i : & W \otimes A & \to & 2^{W \otimes A} \\ & \langle w, \delta \rangle & \mapsto & f(w) \times g(\delta) \end{array}$

The application, via product update, of an action model on a state model is supposed to represent the (more or less epistemically determined) action represented by the action model, and the effect it has on the epistemic status of the agents in the state model. Clearly the public announcement action discussed in section 2.3.2 is a special case of an action model: the public announcement of ϕ is given by the action model consisting of a single point q where each agent at q sees q, and q's precondition is ϕ.

Definition 2.3.10 (DEL model):
An n-agent DEL model is a pair $\langle \mathcal{S}, \mathcal{A} \rangle$ where $\mathcal{S}$ is an n-agent state model and $\mathcal{A}$ is an n-agent action model.

There is a complete logic for such models[4], including a modality for update; see for example [4]. We refer to this logic as dynamic epistemic logic, or DEL.

It is easily seen that PAL is a special case of DEL: An announcement to $\mathcal{G}$ that ϕ is an action model with two points, as depicted in figure 2.1.

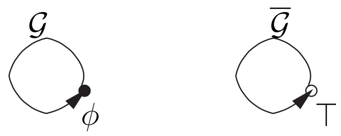


Figure 2.1: Action model representing the Public Announcement to $\mathcal{G}$ that ϕ

Now we can quote the other half of what we will call the 'epistemic representation thesis' (i.e. the following plus the epistemic state representation thesis):

[4]At least: For the case of finitely-branching action models; otherwise an infinitary language would be required.

Thesis 2 (Epistemic Action Representation):
"Let σ be a *social "action"* involving and affecting the knowledge (beliefs, common knowledge) of agents. This naturally induces a *change of situation*; i.e., an operation o taking situations s into situations $o(s)$. Assume that o is presented by assertions concerning knowledge, beliefs and common knowledge facts about s and $o(s)$, and that o is completely determined by these assertions. Then

(a) We may associate to the action σ a mathematical model Σ which we call an [$n-$ *agent*] action model. (Σ is also an [n agent relational] model .[5]) The point again is that all the intuitive features of, and judgments about, σ correspond to formal properties of Σ.

(b) There is an operation $\otimes$ taking a state model S and an action model Σ and returning a new state model $S \otimes \Sigma$. So each Σ induces an *update operation* O on state models: $O(S) = S \otimes \Sigma$.

(c) The update O is a faithful model of the situation change o, in the sense that for all s: if s corresponds to S as in [the epistemic state representation thesis], then again $o(s)$ corresponds to $O(S)$ in the same way; i.e. all intuitive judgements concerning $o(s)$ correspond to formal assertions concerning $O(S)$, and vice-versa." ([3], p. 167)

DEL, like PAL, only allows the modeling of belief *expansion*: In both cases, if an agent believes that p and then learns that $\neg p$, the agent will be in a very unfortunate epistemic state, in which everything is believed. Thus it provides an excellent framework for reasoning about knowledge[6] and its dynamics. Sadrzadeh's statement that DEL "can easily deal with more interesting and closer to real life versions of these puzzles where children perform secret actions such as cheating and lying" ([26], p. 5) is, while perhaps not false, nonetheless misleading. There is indeed nothing problematic about modeling such situations up to a point, namely the point in time where an agent realises that one of her beliefs is wrong, for example discovering the cheating. In such cases the agent will have no capacity to form any coherent beliefs. – There is no scope for *revision.* Thus we can already see a flaw in Baltag and Moss' epistemic representation thesis, with respect to the class of mathematical structures that they propose as candidates to satisfy it: This sort of relational structure is not adequate for discussing *beliefs*, which they do include in the scope of their thesis. This is because beliefs can be revised, whereas DEL does not provide the facility to model this. (The thesis can still be defended with respect to the class of models proposed if 'justifiable belief' means factive or veridical belief, i.e. entailing truth. We will talk in what follows about a 'doxastic version' of the

[5]This is not strictly true in the case of DEL: Action models are n-agent *frames*, but do not in general have a full valuation; cf. 2.3.3.

[6]In fact, although this is peripheral to our present concerns, we note that calling what is modeled "knowledge" is presumptive; some (including Hintikka) have qualified the epistemic states that are modeled by relational frames and, by extension, by DEL models, as "*implicit* knowledge" states.

representation thesis, which we mean specifically to exclude the requirement for such strong justification.)

We will now return to belief revision, and present a formal representation theorem, which provides an enriched semantics for belief states. Later, in §4, we will present ways of blending relational structures with that enriched semantics, in which the 'fine-grainedness' of relational semantics, the importance of points' relations to other points, will necessarily return. Although no class of structures is proposed to satisfy the epistemic representation thesis, it should be borne in mind, as it will partly guide the progression of ideas in this chapter as well as later, and will occasionally be referred back to.

2.3.4 Sphere Semantics

A number of representation theorems have been developed for the AGM postulates. One of them appeared in [1], but remains syntactical. Grove [16] gives a representation theorem which we can more easily relate to relational models.[7] That theorem uses ideas of Lewis [20], who discusses orderings over interpretations in the context of his analysis of counterfactual conditionals. Grove's approach can be retroactively justified by remarking that "If a were to believe that ϕ then a would believe that ψ", a sentence which encodes the essential information about a's disposition for theory change, is a counterfactual conditional of exactly the type with which Lewis was concerned. We will look at logics which can express such conditions (syntactically expressed as $\phi \Box\!\!\Rightarrow \psi$: 'If ϕ were to hold then ψ would hold') starting at the end of the next chapter, and in more detail in chapter 4.

Again, as we did with our presentation of relational semantics, we will adapt the presentation from Grove's paper in order to suit our exposition. Consider a non-empty set W (whose elements we shall again call 'points'). A pre-order over W is a relation $\preccurlyeq \subseteq W \times W$ that is reflexive ($\forall x \in W, x \preccurlyeq x$) and transitive ($\forall \langle x, y, z \rangle \in W^3, ((x \preccurlyeq y$ and $y \preccurlyeq z) \Rightarrow y \preccurlyeq z))$. We call a relation *total* iff $\forall \langle x, y \rangle \in W^2, x \preccurlyeq y$ or $y \preccurlyeq x$. We use the function *min*, given as follows:

Definition 2.3.11 (*min* (Selection)):

$$\begin{array}{llll} min: & 2^W \times 2^{W \times W} & \rightarrow & 2^W \\ & \langle Y, \preccurlyeq \rangle & \mapsto & \{x \in Y | \forall y \in Y, x \preccurlyeq y\} \end{array}$$

That is: $min(X, \preccurlyeq)$ is the set of $\preccurlyeq$-minimal points that are in X.

[7] Although the way we present that theorem does not bear much resemblance to the original presentation, it is on reflection seen to be equivalent; I consider that Grove's original approach *can* be called semantic, contrary to the view expressed by Katsuno and Mendelzon [18], who consider that they have come up with "a semantic counterpart to [Grove's] system of spheres" (op. cit. p.281).

Since we are restricting ourselves here to propositional logic, we present a restricted version of Grove's representation theorem; note that the original theorem is much more general, concerning as it does a wider class of logics. We need some more definitions before presenting that theorem in our setting.

Preliminary Definitions

Definition 2.3.12 (Well-Behaved Relation):
We say that $\preccurlyeq \subseteq W \times W$ is *well-behaved* iff $\varnothing \neq X \subseteq W \Rightarrow min(X, \preccurlyeq) \neq \varnothing$.

Definition 2.3.13 (Anonymous Belief State[8]):
An *anonymous belief state* over W is a well-behaved total pre-order over W. We denote the set of anonymous belief states over W by $\mathcal{O}_W$

An anonymous belief state is intended to represent an agent's preferences for certain points (possibilities) over others. The significant result, that we are coming to after these definitions, is that the existence of a consistent such preference ordering over a certain set of points coincides with making rational revisions, where 'rational' means revisions that concord with the AGM postulates. The set of points over which the preference exists is given in the following model:

Definition 2.3.14 (Canonical One-Shot Model):
$\langle W, \preccurlyeq, V\rangle$ is a *canonical one-shot model* for the language $\mathcal{L}$ over the alphabet Φ of propositional letters iff: $\forall \Psi \subseteq \Phi, \exists w \in W$ such that $V(w) = \Psi$; and $\preccurlyeq$ is an anonymous belief state.

To make the link with the AGM conception of a theory as a deductively closed set of formulae, there is a natural way to make points in the model correspond to complete theories, and to extend this to make sets of points correspond with theories:

Definition 2.3.15 (Theory):
The *theory of a point* w, $Th(w) = (V(w) \cup \{\neg p | p \notin V(w)\})^{\vdash}$. The *theory of a set of points* X, $TH(X) = \cap_{x \in X} Th(x)$.

The adjoint of a theory is the truth set corresponding to it:

Definition 2.3.16 (Truth Set):
The *truth set of a formula* ϕ, $[\![\phi]\!] = \{w \in W | \phi \in Th(w)\}$

Definition 2.3.17 (System of Spheres):
A *system of spheres centered on* $\Gamma \subseteq \Sigma_{\mathcal{L}}$ is a canonical one-shot model $\langle W, \preccurlyeq, V\rangle$ such

[8] The name is from [22]; see §3.2.

that $TH(min(W, \preccurlyeq)) = \Gamma$.

That is, a system of spheres centered on Γ represents the following information about an agent: Her beliefs 'as things stand', Γ, but also her dispositions to revise those beliefs on acquiring new information; those dispositions are given in the rest of the structure of $\preccurlyeq$.

Definition 2.3.18:
The *belief state* determined by Γ and the belief revision function $\dotplus$, is the function $bst^{\dotplus}_{\Gamma} : \mathcal{L} \to \Sigma_{\mathcal{L}}$ such that $bst^{\dotplus}_{\Gamma}(\phi) = \Gamma \dotplus \phi$. (Note that $bst^{\dotplus}_{\Gamma}(\top) = \Gamma$ as long as $\Gamma \neq \mathcal{L}$.)

Definition 2.3.19 (Centred Belief State):
Given a belief state b, we say that b is *centered on* Γ iff $b(\top) = \Gamma$.

Definition 2.3.20 (Respecting AGM):
A belief state b *respects AGM* iff the following versions of ($\dotplus 1 \ldots \dotplus 6$) hold:

($b1$) $b(\phi) \vdash \phi$

($b2$) $b(\top) \nvdash \neg\phi \Rightarrow b(\phi) = (b(\top) \cup \phi)^{\vdash}$

($b3$) $\nvdash \neg\phi \Rightarrow b(\phi) \nvdash \bot$

($b4$) $\vdash \phi \equiv \psi \Rightarrow b(\phi) = b(\psi)$

($b5$) $b(\phi \wedge \psi) \subseteq ((b(\phi)) \cup \{\psi\})^{\vdash}$

($b6$) $b(\phi) \nvdash \neg\psi \Rightarrow (b(\phi) \cup \{\psi\})^{\vdash} \subseteq b(\phi \wedge \psi)$

Definition 2.3.21:
The belief state generated by a system of spheres $\mathfrak{S} = \langle W, \preccurlyeq, V \rangle$ is the function $bst_{\mathfrak{S}} : \mathcal{L} \to \Sigma_{\mathcal{L}}$ such that $bst_{\mathfrak{S}}(\phi) = TH(min([\![\phi]\!], \preccurlyeq))$.

Sphere Representation

Theorems 1 and 2 in [16] can now be phrased in our terms as follows.

Theorem 2.3.1:
If a belief state b centered on Γ respects AGM then there exists a system of spheres $\mathfrak{S}$ centered on Γ such that $b = bst_{\mathfrak{S}}$.

Theorem 2.3.2:
The belief state generated by a system of spheres centered on $\Gamma \neq \mathcal{L}$ respects AGM.

This representation theorem is certainly a step towards having a model which would satisfy the doxastic version of Baltag and Moss' thesis: We have an accessible mathematical model for representing a single agent rationally accepting a new piece of information. However, it is clearly very limited. There is no facility for modeling several agents or for repeated revisions. That will come in the next chapters, as we consider natural extensions of these kinds of structure.

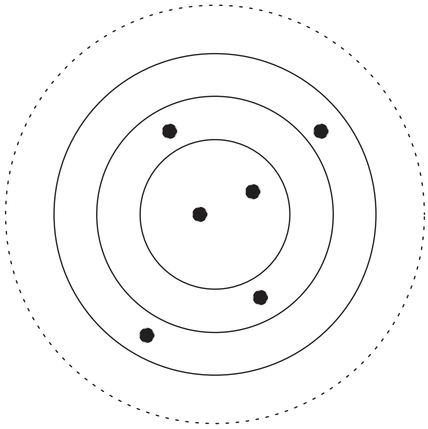


Figure 2.2: A belief onion

Anonymous belief states can be represented as 'onions' in diagrams like figure 2.2. It is for this reason that we call such models 'sphere' models: Each ring of the onion represents a set of points (a sphere), namely those points which are smaller than some point according to the pre-order. We will use these diagrams in the next chapter to consider how several of them can be combined, or 'fused', to form on. After that we will consider richer structures, in which an anonymous belief state is associated with each point, so we would have a diagram looking more like rain on an Amsterdam canal than an onion.

3 Merging Beliefs

3.1 Introduction

Some of the problems raised in the last chapter concerned belief revision amongst *groups* of agents. The fact that there were several agents was important because they had to revise their beliefs about each others' beliefs, and this is obviously more complex than in the single-agent non-introspective case. In this chapter, we do not concern ourselves with those higher-order belief issues, but rather look at how beliefs of groups of agents can be merged to form a single, collective belief state. We start off (§3.2) by presenting one account of belief 'fusion' due to [22]. We then (§3.3) show how it relates to the AGM postulates for revision. In [22], the authors show how fusion can determine a specific form of iterated belief revision. We show here that one need only what is in [22] called 'belief refinement' in order to determine that form of belief revision. We sketch a modal logic which seems to me the most natural one for describing anonymous belief states, and we show that it can express AGM-style belief revision, and belief refinement.

3.2 Belief Fusion

3.2.1 Presentation

Maynard-Reid II and Shoham [22] introduced the name *anonymous belief states* for total well-behaved transitive pre-orders over a set. These are the same objects that play the role of the implicit agent's belief state in Grove-style canonical one-shot models (see 2.3.4). Here we will use $\mathcal{B}_W$ to denote the set of anonymous belief states over W. We will use $\preccurlyeq$ and slight typographical variants to denote anonymous belief states. Then $\prec$ will denote the irreflexive restriction of $\preccurlyeq$ (i.e. $x \prec y \Leftrightarrow (x \preccurlyeq y \& y \not\preccurlyeq x)$), and $\sim$ will denote $\preccurlyeq$-equivalence (i.e. $x \sim y \Leftrightarrow x \preccurlyeq y \wedge y \preccurlyeq x$).

Belief Refinement

[22] describes two ways to combine such belief states in order to extract the most information from them, where certain belief states are considered to have a higher priority

(to be more reliable) than others. The first is called *refinement*: If $\preccurlyeq_A$ and $\preccurlyeq_B$ are belief states, then refining $\preccurlyeq_A$ by $\preccurlyeq_B$ yields $\preccurlyeq_A \veebar \preccurlyeq_B$, where

Definition 3.2.1 (Belief Refinement):
$w_1(\preccurlyeq_A \veebar \preccurlyeq_B)w_2 \Leftrightarrow_{df} (w_1 \prec_A w_2 \textit{ or } (w_1 \sim_A w_2 \textit{ and } w_1 \preccurlyeq_B w_2))$

Belief refinement is a kind of lexicographic combination of the orders. It is a natural enough operation, and one that will appear, albeit in slightly different forms, several times in this thesis.

[22] point out that belief refinement is a forgetful operator, in the sense that "the standard belief state is not rich enough to represent the source of each information item" (op. cit., p. 186). That is: suppose there are three belief states $\preccurlyeq_A$, $\preccurlyeq_B$ and $\preccurlyeq_C$ which are to be combined, where $\preccurlyeq_C$ is to take priority over (i.e. be more trustworthy than) $\preccurlyeq_B$ and $\preccurlyeq_A$, and $\preccurlyeq_B$ to take priority over $\preccurlyeq_A$. [22] makes the simple observation that the order in which the refinement operation is carried out is the only way to control this priority. That is, $\preccurlyeq_C$ must be refined by $\preccurlyeq_B$, and then $\preccurlyeq_A$ must refine the outcome: The result should be $(\preccurlyeq_C \veebar \preccurlyeq_B) \veebar \preccurlyeq_A$, and the operation is not commutative. To take the most simple example:

$$a \sim_C b$$
$$a \prec_B b$$
$$b \prec_A a$$

These give:

$$a \prec_{C \veebar B} b, \Rightarrow a \prec_{(C \veebar B) \veebar A} b$$
$$b \prec_{C \veebar A} a, \Rightarrow b \prec_{(C \veebar A) \veebar B} a.$$

Belief Fusion

In order to make a commutative operation, [22] proposes a different operation which does not 'forget' where the information is from. To do this, they introduce *pedigreed belief states*, which are functions from pairs of elements of W into a set of *sources*, where a *source* is just an object s to which is associated an anonymous belief state $\preccurlyeq_s$. So if $\mathcal{S}$ is the set of sources

Definition 3.2.2 (Pedigreed Belief State):
A *pedigreed belief state* is$_{df}$ a function

$$\Phi : W \times W \to \wp\mathcal{S}$$

(Notice the connection between pedigreed and anonymous belief states: the latter can be thought of as special cases of the former, where only one source is considered.) Given $S \subseteq \mathcal{S}$, [22] gives the following

Definition 3.2.3 (Pedigreed Belief State Induced by S):
Φ_S, the *pedigreed belief state induced by* S, is$_{df}$

$$\begin{array}{rrcl} \Phi_S : & W \times W & \rightarrow & \wp S \\ & \langle w_1, w_2 \rangle & \mapsto & \{ s \in S | w_1 \prec_S w_2 \} \end{array}$$

I.e., for any pair of members of W, a pedigreed state says which sources (if any) strictly prefer the one over the other.

Then a new fusion operator, $\amalg$, is introduced:

Definition 3.2.4 (Belief Fusion):
The *fusion* of the pedigreed belief states induced by S and S' is$_{df}$ the pedigreed belief state induced by $S \cup S'$. I.e.

$$\Phi_S \amalg \Phi_{S'} =_{df} \Phi_{S \cup S'}$$

3.2.2 Discussion

Clearly $\amalg$ will behave like $\cup$, so that it is commutative and associative. [22] point out that, given a strict rank $\sqsupset$ over S, one can define an anonymous belief state from S, which prefers w_1 over w_2 just when if there is a source $s \in S$ such that $w_1 \prec_s w_2$, then for any $s' \in S$, if $w_2 \prec_{s'} w_1$ then $s \sqsupset s'$. Thus they effectively define a function

$$\begin{array}{rrcl} A : & \wp(\mathcal{S} \times \mathcal{S}) & \rightarrow & \wp(W \times W) \\ & R & \mapsto & \{ \langle w_1, w_2 \rangle | \forall s (w_1 \prec_s w_2 \Rightarrow \forall s' (w_2 \prec_{s'} w_1 \Rightarrow sRs')) \} \end{array}$$

They consider only the case where R is a strict rank over a subset of $\mathcal{S}$. In such cases, it is easy to see (and also their lemma 1) that

Remark 3.2.1:
where $\sqsupset$ is a strict order over distinct sources $s_n \sqsupset s_{n-1} \ldots \sqsupset s_1$,

$$A(\sqsupset) = ((\preccurlyeq_{s_n} \veebar \preccurlyeq_{s_{n-1}}) \veebar \ldots \preccurlyeq_{s_1}).$$

As we noted, anonymous belief states are appropriately named, as the origin of information is lost. To generate an anonymous belief state from an enriched pedigreed belief state, simply consider what the top source (the 'expert') says, and if it is indifferent, consider the next source, and so on (down to the 'novice').

The enriched definition, then, only has value inasmuch as it would be used to consider sequential refinements, in some kind of dynamic setting: Otherwise, we simply say which sources we are combining, and refine in order of rank. It would have been more illuminating of Shoham and Maynard-Reid II to make this point explicit when they point out that five existing proposals[1] for iterated belief revision do not respect associativity, and indeed lead to contradictory results depending on order of revision. The strict ranking over sources required by their "fusion" is equivalent to requiring that one revise by 'listening' first to highest-ranked sources, then to next-highest and so on.

3.3 Relation to the AGM tradition

3.3.1 A Revision is a refinement

[21] remarks that "AGM revision is a simply a projection of belief fusion between conflict-free agents where one ignores all but the belief set of the expert's belief state and all but the belief set of the resulting belief state". One-shot AGM revision can be seen to be a special case of belief *refinement*.

The elements of the one-shot canonical models (2.3.14), just as in Grove's original systems of spheres, were maximally consistent sets of formulae, but the idea can be retained regardless of what we will want the elements of W to be.[2]

Maynard-Reid II and Shoham write that a "*revision* [function based on the AGM postulates] is a uniquely defined operation that takes as its first argument not a mere belief set, but a full belief state" ([22] p. 192, their emphasis). Another way to see it is the one we adopted in §2.2.1: that a particular AGM revision operator $\Gamma \dotplus -$ is a function from $\mathcal{L}$ into $2^{\mathcal{L}}$.

Definition 3.3.1 (One-Shot Belief Revision Function):
For $\Gamma \subset \mathcal{L}$, f_Γ is a *one-shot belief revision function* from Γ iff$_{df}$ it is of the form

$$\begin{array}{rccc} f_\Gamma : & \mathcal{L} & \rightarrow & 2^{\mathcal{L}} \\ & \phi & \mapsto & \Gamma \dotplus \phi, \end{array}$$

where $\Gamma\dotplus$ respects the AGM postulates (§2.2.1).

Now assume that we are given some one-shot belief revision function $\Gamma\dotplus$. Define the following anonymous belief state over $W = \{\Sigma | \forall \phi \in \mathcal{L}, \phi \in \Sigma \Leftrightarrow \neg\phi \in \Sigma$ and Σ is

[1] Viz. those of Boutilier, Darwiche and Pearl ([10]), Lehman, Spohn ([32]) and Williams.

[2] This restriction is enough to prompt Katsuno and Mendelzon to talk about "a semantic counterpart to the system of spheres" ([18], p. 281), which is almost enough to make one wonder whether they momentarily forgot the close connection between maximally consistent sets of sentences and interpretations, which goes back at least to Henkin.

consistent}

$$w_1 \preccurlyeq_\phi w_2 \Leftrightarrow \phi \in w_2 \Rightarrow \phi \in w_1.$$

This is the ordering the expert places on the points, or rather to look at it another way: a fragment of the expert's view that the novice has acquired via the observation (= statement by the expert) that ϕ holds. Theorem 2.3.1 gives us the novice's view of things: $\preccurlyeq_\Gamma \in \mathcal{B}_W$ such that $min(W, \preccurlyeq_\phi \veebar \preccurlyeq_\Gamma) = \Gamma \dotplus \phi$.

In a more general case where we have a satisfaction relationship $\models$ between points and formulae, we can give the following definition of the 'expert's view of things' (i.e. the binary anonymous belief state induced by the piece of incoming information ϕ):

$$(\star) \quad w \preccurlyeq_\phi w' \Leftrightarrow w' \models \phi \Rightarrow w \models \phi$$

We write $[\![\phi]\!]$ for the set of points w where $w \models \phi$, and $[\![\Sigma]\!]$ for $\bigcup_{\phi \in \Sigma} [\![\phi]\!]$ when Σ is a set of formulae. We say that an anonymous belief state is 'centered' on a set of formulae Γ just when $min(\preccurlyeq) = [\![\Gamma]\!]$. Then we can formulate the following proposition:

Proposition 3.3.1:
If $\preccurlyeq_\Gamma$ is an anonymous belief state centered on Γ, then $\dotplus$, defined thus:

$$\begin{array}{rccl} \dotplus : & \mathcal{L} & \rightarrow & \Sigma_\mathcal{L} \\ & \phi & \mapsto & min(\preccurlyeq_\phi \veebar \preccurlyeq_\Gamma) \end{array}$$

is an *AGM* revision operator for Γ (cf. §2.2.1).

Proof. This proposition is (very differently phrased but) in effect proved in [22], for the case of fusion, which is just ordered refinement. ☻

3.3.2 Iterating revisions

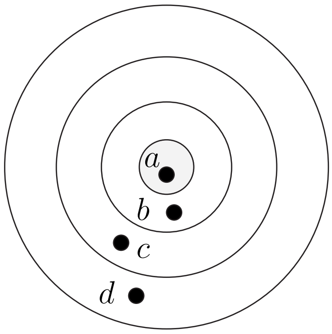
Shoham and Maynard-Reid II talk about an "asymmetry" in existing theories of belief revision: that its output is a belief set, whereas in a sense[3] its input is a belief state. In their setup this is not the case: the output is of the same kind as the input, so that there is an obvious way to iterate revisions. We should think then of the belief revision operation as moving from belief states to belief states, given a formula.

It is instructive to see what this can bring to bear on other literature on iterated belief revision. [10] argue that the AGM postulates themselves are too weak for iterated belief revision, giving the example we depict here in figure 3.1[4].

[3] viz. the sense that the output is not determined by a belief set alone

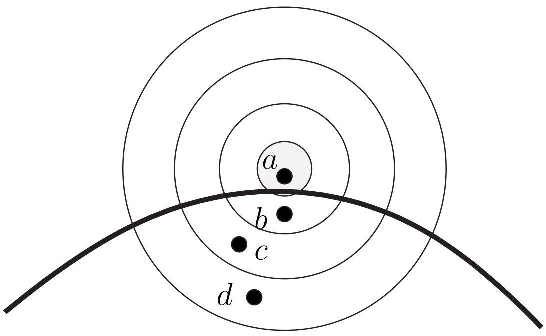
[4] This is, modulo our presentation and insights, a graphical representation of example 6, to be found in appendix A of [10].

Figure 3.1: The belief onion for Γ



Here $[\![p]\!] = \{a, b\}$ and $[\![q]\!] = \{a, c\}$. Darwiche and Pearl consider the following ordering that might obtain after revision by $\neg(p \wedge q)$: $b \prec_! a \sim_! d \prec_! c$, remarking that it is compatible with the AGM postulates. Yet if this were the ordering after the revision by $\neg(p \wedge q)$, we would have $\Gamma \dot{+} (\neg p) \models \neg p \wedge q$[5], but $(\Gamma \dot{+} \neg(p \wedge q)) \dot{+} \neg p \models \neg p \wedge \neg q$. This counterintuitive result should not surprise us, who have seen that the AGM way of thinking, which is to use sets of beliefs, at least as output from the revision, does not give enough constraint to the process to make sense of iterated revision. Notice that on the Shoham and Maynard-Reid II approach, the ordering $\prec_!$ does *not* result from refinement. Rather, the refinement operator provides a way to determine uniquely the resulting ordering, which will be as in figure 3.2.

Figure 3.2: Γ about to be revised by $\neg(p \wedge q)$

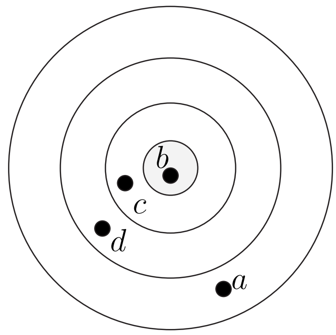


Here the $\neg(p \wedge q)$-points form the core of the expert's anonymous belief state, which is then $b \sim c \sim d \prec a$, leading to the refined belief state represented in 3.3.

Now if this were to be revised by $\neg p$, as in Darwiche and Pearl's example, we get the

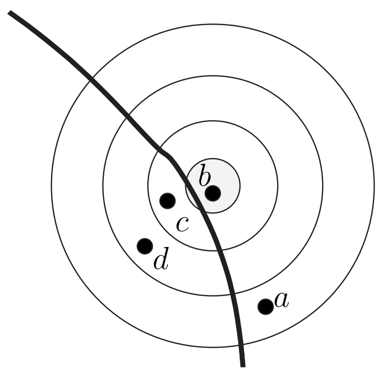
[5]We write $\phi \models \psi$ to mean $[\![\phi]\!] \subseteq [\![\psi]\!]$. We are sloppy with things like domains and models in this chapter until we introduce a more fully formal logical approach in §3.4.

Figure 3.3: The belief onion for $\Gamma \dot{+} \neg(p \wedge q)$



anonymous belief state equivalent to the belief onion represented in figure 3.4.

Figure 3.4: $\Gamma \dot{+} \neg(p \wedge q)$ about to be revised by $\neg p$



I.e. the expert's anonymous belief state is $c \sim d \prec b \sim a$. This in turn yields a (unique) refinement, viz. the anonymous belief state equivalent to the belief onion represented in figure 3.5, i.e. $c \prec d \prec b \prec a$.

As we said, this kind of iteration of revision is a natural one and will occur. There are even more pictures of it 'in action' in §4.2.2.

Darwiche and Pearl introduce new postulates for iterated revision:

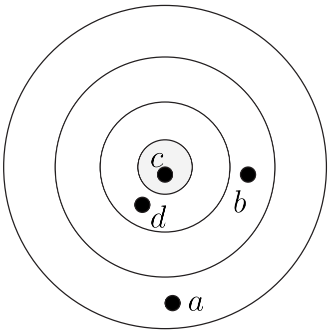
(DP1) $\phi \models \psi \Rightarrow (\Gamma \dot{+} \psi) \dot{+} \phi \equiv \Gamma \dot{+} \phi$

(DP2) $\phi \models \neg\psi \Rightarrow (\Gamma \dot{+} \psi) \dot{+} \phi \equiv \Gamma \dot{+} \phi$

(DP3) $\Gamma \dot{+} \phi \models \psi \Rightarrow (\Gamma \dot{+} \psi) \dot{+} \phi \models \psi$

(DP4) $\Gamma \dot{+} \phi \not\models \neg\psi \Rightarrow (\Gamma \dot{+} \psi) \dot{+} \phi \not\models \neg\psi$

Figure 3.5: $(\Gamma \dotplus \neg(p \wedge q)) \dotplus \neg p$



We shed some light on these postulates by proving that they hold for belief refinement, re-written following remark 3.2.1, i.e. where the belief set of $(\Gamma \dotplus \phi_0) \ldots \dotplus \phi_n$ is the theory of $min(W, \preccurlyeq_{\psi_n} \veebar \ldots (\preccurlyeq_{\phi_0} \veebar \preccurlyeq_\Gamma))$.

Proposition 3.3.2:
Taking $\Gamma \equiv \Gamma'$ to mean that the belief *set* of Γ is logically equivalent to the belief set of Γ', the Darwiche and Pearl postulates for iterated revision hold with respect to belief refinement, for revision by non-absurd propositions.[6]

Proof. We must show that each of the above postulates holds when the revision is iterated according to belief refinement.

Lemma 3.3.1 (Success Lemma):
The fact that $\phi \not\models \bot$ means that for any anonymous belief state $\preccurlyeq$, if $x \in min(\preccurlyeq_\phi \veebar \preccurlyeq)$ then $x \models \phi$.

Proof. Take any $x \in min(\preccurlyeq_\phi \veebar \preccurlyeq)$. Then for any y in the domain, $x(\preccurlyeq_\phi \veebar \preccurlyeq)y$. So in particular, for one of the y's such that $y \models \phi$ (remember that $\phi \not\models \bot$), $x(\preccurlyeq_\phi \veebar \preccurlyeq)y$. So either $x \prec_\phi y$, or $x \sim_\phi y$ (and $x \preccurlyeq_\Gamma y$). The former would require that $y \not\models \phi$, which is *contra* our assumption; the second entails that $y \models \phi \Leftrightarrow x \models \phi$, so indeed $x \models \phi$. ☻

.

That lemma will be useful in what follows.

For something approaching readability, we will in this proof write xAy for $x[\preccurlyeq_\phi \veebar(\preccurlyeq_\psi \veebar \preccurlyeq_\Gamma)]y$, xBy for $x[\preccurlyeq_\phi \veebar \preccurlyeq_\Gamma]y$, and xCy for $x[\preccurlyeq_\psi \veebar \preccurlyeq_\Gamma]y$.

[6] I.e. for any ϕ such that $\phi \not\models \bot$.

(DP1) Suppose that $\phi \vDash \psi$. We must show that $min(A) = min(B)$. There are thus two inclusions:

$\subseteq$: Then take any $b \in min(A)$. Then take any z in the domain; we must show that bBz. Our hypothesis that $b \in min(A)$ tells us that bAz. This gives two possibilities:

1. $b \prec_\phi z$. Then we have bBz immediately.
2. $b \sim_\phi z$ and bCz. This again gives rise to two possibilities:
   a) $b \prec_\psi z$. Then $b \vDash \psi$ and $z \nvDash \psi$. So because $\phi \vDash \psi$, $z \nvDash \phi$. So because $b \sim_\phi z$, $z \nvdash \phi$. Now either:
      i. $b \vdash \phi$, in which case $b \prec_\phi z$, so bBz; OR:
      ii. $b \nvdash \phi$. But the success lemma tells us that this is not possible.
   b) $b \sim_\psi z$ and $b \preccurlyeq_\Gamma z$. Since $b \sim_\phi z$, we have bBz.

$\supseteq$: Take any $b \in min(B)$. So take any z; we must show that bAz. Our hypothesis again gives us two initial possibilities:

1. $b \prec_\phi z$; then bAz.
2. $b \sim_\phi z$ and $b \preccurlyeq_\Gamma z$. So $z \vDash \phi$, and since $\phi \vDash \psi$, $b \vDash \psi$ and $z \vDash \psi$. So $b \sim_\psi z$. So indeed bAz.

(DP2) Assume that $\phi \vDash \neg\psi$. Now again we must show that $min(A) = min(B)$:

$\subseteq$: Take any $b \in min(A)$. Take a z such that bAz. We must show that bBz. Because bAz, there are two possibilities:

1. $b \prec_\phi z$. In this case bBz.
2. $b \sim_\phi z$ and bCz. By the success lemma $b \vDash \phi$, so $b \vDash \neg\psi$. Then $b \sim_\psi z$ and $b \preccurlyeq_\Gamma z$. So bBz.

$\supseteq$: Take any $a \in min(B)$. Take a z with bBz. Two possibilities:

1. $b \prec_\phi z$; then bAz.
2. $b \sim_\phi z$ and $z \preccurlyeq_\Gamma b$. Now since $\phi \vDash \psi$ and $b \vDash \phi$ and $b \sim_\phi z$, then $b \sim_\psi z$. So bAz.

(DP3) Take as a hypothesis that $min(B) \subseteq [\![\psi]\!]$. Now $b \in min(B)$ iff $\forall z, b \prec_\phi z$ *or* $(b \sim_\phi z \ \& \ b \preccurlyeq_\Gamma z)$.

We must show that $min(A) \subseteq [\![\psi]\!]$. Suppose towards a contradiction that this is not so. That is, that there is an $e \in min(A)$ such that $e \nvDash \psi$. Now because $e \in min(A)$, for every z one of the three following situations holds:

1. $e \prec_\phi z$;
2. $e \sim_\phi z \ \& \ e \prec_\psi z$;
3. $e \sim_\phi z \ \& \ e \sim_\psi z \ \& \ e \preccurlyeq_\Gamma$.

However, notice that by our supposition that $e \models \psi$, the second of these situations is impossible. But then $e \in min(B)$, so by our hypothesis that $min(B) \subseteq [\![\psi]\!]$, we have then contradicted that supposition, $e \models \psi$.

(DP4) Assume that $\Gamma \dotplus \phi \not\models \neg\psi$, we must show that $(\Gamma \dotplus \psi) \dotplus \phi \not\models \neg\psi$. Our assumption is: $\exists b \in min(B) : b \not\models \psi$. By the success lemma we know that $b \models \phi$. So we know that $\forall z(z \not\models \phi \Rightarrow b \preccurlyeq_\Gamma z)$.

We will show that $b \in min(A)$, completing the proof. That is, we must show that given any z, $z \models \phi \;\Rightarrow\; (a \prec_\psi z \; or \; (a \sim_\psi z \;\&\; a \preccurlyeq_\Gamma z))$. Then we must show that if $(\star)$ $a \preccurlyeq_\Gamma z$ then either $a \prec_\psi z$ or $a \sim_\psi z \wedge a \preccurlyeq_\Gamma z$. Since $a \models \psi$, this means: either $z \models \neg\psi$ or $z \models \psi$ & $a \preccurlyeq_\Gamma z$. This is entailed by $(\star)$.

☻

Belief refinement is a strategy for iterated revision that will appear again in the next chapter. We will also propose different postulates to characterise this and different kinds of iterated belief revision (§4.2.2).

3.4 Logic for Anonymous Belief States

3.4.1 Logical Syntax

We have only so far an entirely semantic characterisation of one kind of belief 'fusion' or 'merging'. Thus two questions naturally arise. Firstly, what about syntax? And secondly, what about other ways of combining belief states?

To illustrate part of the interest of the first question, consider how what is supposedly being represented by the 'fusion' described above might take place. Information is usually considered to arrive in packets: a statement or observation claims to reveal some aspect of reality. Furthermore, information from less reliable agents is of most use not necessarily in arbitrating between elements of W which the most reliable agent does not distinguish, but perhaps more importantly when we lack information from the most reliable agent: When *as far as we know* that agent does not distinguish between points.

As described in section 2.1, we can model the gradual acquisition of information by considering formulae of a language, which can be true or false at points. So we might start with a collection of proposition letters Φ, the standard Boolean operators, and associate a valuation function $V : \Phi \to 2^W$ with the belief states in question. Then consider how much information about a source's ordering can be obtained if they emit formulae of this language. In order for anything like the semantic merging we have described to take place, these formulae must be able to express more than just the belief set of the agent. That is, the agent must be able to express more than just some proposition ϕ representing the fact that all the points that are minimal for her ordering are ϕ points. If a source could state only such propositions, it would not be possible to acquire much information about the richer structure of her beliefs. But it is precisely this structure that is of interest in Shoham and Maynard-Reid II's modeling, because it might be the case that there is a source s who has the following belief state: $u \prec v \prec w$, and a more reliable source who has the following state: $v \sim w \prec u$. Then no useful information would be available from s if she can only state propositional formulae that she believes. What we want to do is to say to s, "suppose we know u not to be an accurate representation of the actual world; what then?", at which point she would be able to reveal to us facts about v. In other words, we want agents to be able to make counterfactual conditional statements. That is, an agent i should be able to make a (*counterfactual*) statement like 'Well, if ϕ has to be the case, then ψ'.

That is exactly the format of the logical language used in §4.3. A slightly different approach will be presented in §4.2. Before that, in the remainder of this chapter, we will sketch a natural logical syntax that can be interpreted over these structures.

3.4.2 A Sketch of Syntax and Semantics

The logic that we sketch in this section we call *LAB*: A Logic for Anonymous Belief states

We can notice that the pre-order relation can be 're-packaged' as a unary modality: Given any $\preccurlyeq \in \mathcal{B}_W$ (which can be denoted $\langle W, \preccurlyeq \rangle$) we will define $f^<$ as follows:

$$\begin{array}{rccl} f^< : & W & \to & W \\ & w & \mapsto & \{v \in W | w \preccurlyeq v \ \& \ v \not\preccurlyeq w\} \end{array}$$

I.e. $f^<(w)$ gives the set of strictly 'better' points.

Then we define $f^<$ to be the converse of $f^>$:

$$w \in f^>(v) \Leftrightarrow_{df} v \in f^<(w)$$

Finally, we will define $f_k(w)$ as the constant function returning W.

Definition 3.4.1 (Anonymous Belief Frame):
Given $\preccurlyeq \in \mathcal{B}_W$, the tuple

$$\langle W, f^<, f^>, f_k \rangle$$

Is *an anonymous belief frame* just when$_{df}$ $f^<$, $f^>$ and f^k are defined as above.

Then, if we equip an anonymous belief frame with a valuation V, we can use the standard modal semantics, using $\Diamond^{<}$, $\Diamond^{>}$ and k as modal operators (and $\Box^{<}$, $\Box^{>}$ and K for their duals) to interpreting the frame's functions, to say a great deal about the structures.

Fact 1:
The following axioms are sound with respect to anonymous belief frames:

($\Box$K) $\Box(p \supset q) \supset (\Box p \supset \Box q)$ for $\Box \in \{\Box^{<}, \Box^{>}, K\}$.

$(K4)$ $K\phi \supset KK\phi$

$(K5)$ $k\phi \supset Kk\phi$

($\Box^{>}$L) $\Box^{>}(\Box^{>}p \supset p) \supset \Box^{>}p$

We can consider frames that combine several agents' belief states, in order to use this logical language to talk about belief refinement:

Definition 3.4.2 (n-agent Anonymous Belief Frame):
Given $\{\preccurlyeq_i \in \mathcal{B}_{W_i}\}_{i \in n}$, the tuple

$$\langle \Omega, \{f_i^{<}, f_i^{>}, f_i^{k}\}_{i \in n} \rangle$$

Is *an n-agent anonymous belief frame* just when$_{df}$

1. for each $i \in n$, $\langle \Omega, f_i^{<} \restriction W_i, f_i^{>} \restriction W_i, f_i^{k} \restriction W_i \rangle$ (the restriction to W_i) is an anonymous belief frame; and
2. $\forall i \in n$, $\forall w \in \Omega - W_i$,
   a) $f_i^{>}(w) = \varnothing$
   b) $f_i^{<}(w) = f_i^{k}(w) = W_i$
3. $\Omega = \bigcup_{i \in n} W_i$.

We say then that $\langle W_i, \{f_i^{<}, f_i^{>}, f_i^{k}\}_{i \in n} \rangle$ is the anonymous belief frame *induced by* $\{\preccurlyeq_i \in \mathcal{B}_{W_i}\}_{i \in n}$.

3.4.3 Expressing Merging

Now we can use this logic to define a number of belief merging operators. Here we will choose refinement: For a simple example we will consider the 2-agent case, and will introduce a new operator $\Diamond^{<}_{1 \veebar 2}$.

Consider the following axiom:

(A$\veebar$) $\Diamond^{<}_{1\veebar 2}\phi \equiv (\Diamond^{<}_{1}\phi \vee (\neg\Diamond^{>}_{1}\phi \wedge \Diamond^{<}_{2}\phi))$

We show that this axiom characterises the operation of belief refinement, in the following sense:

Theorem 3.4.1 (Characterisation of Refinement):
Given a 2-agent anonymous belief frame $\mathcal{F} = \langle \Omega, \{f_i^<, f_i^>, f_i^k\}_{i\in\{1,2\}}\rangle$, the formula (A$\veebar$) is valid on the structure $\langle \Omega, \{f_i^<, f_i^>, f_i^k\}_{i\in\{1,2\}}, f^<_{1\veebar 2}\rangle$ precisely when $f^<_{1\veebar 2}$ is the relation obtained from $\preccurlyeq_1 \veebar \preccurlyeq_2$, where $\mathcal{F}$ is induced by $\{\preccurlyeq_1, \preccurlyeq_2\}$.

Proof. First, we observe an equivalent way of writing the definition of belief refinement, in terms of the strict $\prec$ relations rather than the reflexive $\preccurlyeq$ relations:

Lemma 3.4.1 (Strict Refinement):
Definition 3.2.1 is equivalent to:

$$x(\prec_A \veebar \prec_B)y \Leftrightarrow (x \prec_A y \textit{ or } (y \nprec_A \ \& \ x \prec_B y))$$

Proof. We write $\prec_C$ for $(\prec_A \veebar \prec_B)$:

$$\begin{array}{rcl} x \prec_C y & \Leftrightarrow & (x \preccurlyeq_C y \ \& \ y \not\preccurlyeq_C x) \\ & \Leftrightarrow & ((x \prec_A y \textit{ or } (x \sim_A y \ \& \ x \prec_B y)) \\ & & AND \\ & & (y \nprec_A x \ \& \ (y \nsim_A x \ \& \ y \not\preccurlyeq_B x))) \\ & \Leftrightarrow & (x \preccurlyeq_A y \ \& \ y \nsim_A x) \\ & & OR \\ & & (x \preccurlyeq_B y \ \& \ y \not\preccurlyeq_B x \ \& x \sim_A y) \\ & \Leftrightarrow & x \prec_A y \\ & & OR \\ & & y \nprec_A x \ \& \ x \prec_B y \end{array}$$

☻

Then the proposition to be proved can be re-expressed as follows:

$$\begin{array}{rl} \mathcal{F} \models (A\veebar) & \Leftrightarrow \forall w \in \Omega, f^<_{1\veebar 2}(w) = \{w' | w' \prec_{1\veebar 2} w\} \\ & \Leftrightarrow \forall w \in \Omega, f^<_{1\veebar 2}(w) = \{w' | w' \prec_1 w \textit{ or } (w \nprec_A w' \ \& \ w' \prec_B w)\} \end{array}$$

We will prove the two directions of this bi-implication one at a time:

$\Rightarrow$: Suppose that $\mathcal{F} \vDash$(A$\veebar$). In particular then, for any w, $\mathcal{F}, w \vDash$ (A$\veebar$). Now we must show that $f^{<}_{1\veebar 2}(w) = \{w'|w' \prec_1 w \textit{ or } (w \nprec_A w' \;\&\; w' \prec_B w)\}$. We will show each direction of the inclusion:

$\subseteq$: Take some $w' \in f^{<}_{1\veebar 2}(w)$. Then choose a valuation V such that $V(p) = \{w\}$. Then:

$$
\begin{array}{ll}
 & w \vDash \Diamond^{<}_{1\veebar 2} p \\
\Rightarrow & w \vDash \Diamond^{<}_{1} p \vee (\neg \Diamond^{>}_{1} p \wedge \Diamond^{<}_{2} \phi) \quad \text{(by the validity of (A}\veebar\text{))} \\
\Leftrightarrow & w \vDash \Diamond^{<}_{1} p \\
 & OR \\
 & w \vDash \neg \Diamond^{>}_{1} p \wedge \Diamond^{<}_{2} p \\
\Leftrightarrow & \exists w'' \in f^{<}_{1}(w) : w'' \vDash p \\
 & OR \\
 & \exists w'' \in f^{<}_{2}(w) : w'' \notin f^{>}_{1}(w)) \\
\Leftrightarrow & w' \in f^{<}_{1}(w) \\
 & OR \\
 & w' \in f^{<}_{2}(w) \;\&\; w' \notin f^{>}_{1}(w) \\
\Leftrightarrow & w' \prec_1 (w) \\
 & OR \\
 & w' \prec_2 (w) \;\&\; w \nprec_1 (w) w'
\end{array}
$$

$\supseteq$: Take any w in the domain. Here we must show that for any w' such that (a) $w' \prec_1 w$ or (b) $w' \prec_2 w \;\&\; w \nprec_1 w'$, it holds that $w' \in f^{<}_{1\veebar 2}$. That is: We must show that both (a) and (b) determine this.

(a) Take a w' such that $w' \prec_1 w$. Then choose a valuation V such that $V(p) = \{w'\}$. Now $w \vDash \Diamond^{<}_{1} p$. So by the validity of (A$\veebar$), $w \vDash \Diamond^{<}_{1\veebar 2} p$, so $w' \in f^{<}_{1\veebar 2}(w)$.

(b) Take a w' such that $w' \prec_2 w$ and $w \nprec_1 w'$. Again, choose a valuation such that $V(p) = \{w'\}$. Then $w \vDash \neg \Diamond^{>}_{1} p \wedge \Diamond^{<}_{2} p$, so $w' \notin f^{>}_{1}(w)$ and $w' \in f^{<}_{2}(w)$.

$\Leftarrow$: The easy direction: Suppose that for all w, $f^{<}_{1\veebar 2}(w) = \{w'|w' \prec_1 w \textit{ or } (w \nprec_A w' \;\&\; w' \prec_B w)\}$. Then check that $\mathcal{F} \vDash$(A$\veebar$):

So let $\mathcal{M}$ be a model based on $\mathcal{F}$, and w a member of the domain. Then:

$$\begin{array}{rcl}
\mathcal{M}, w \vDash \Diamond^{<}_{1 \veebar 2}\phi & \Leftrightarrow & \exists w' \in [\![\phi]\!] \cap f^{<}_{1 \veebar 2}(w) \\
& \Leftrightarrow & \exists w' \in [\![\phi]\!] \cap \{w'' | w'' \prec_1 w \ or \ (w \nprec_A w'' \ \& \ w'' \prec_B w)\} \\
& \Leftrightarrow & \exists w' \in [\![\phi]\!] : (w' \prec_1 w \ or \ (w \nprec_A w' \ \& \ w' \prec_B w) \\
& \Leftrightarrow & \exists w' \in [\![\phi]\!] : w' \prec_1 w \\
& & OR \\
& & \exists w' \in [\![\phi]\!] : w \nprec_A w' \ \& \ w' \prec_B w \\
& \Leftrightarrow & \mathcal{M}, w \vDash \Diamond^{<}_{1}\phi \\
& \Leftrightarrow & OR \\
& \Leftrightarrow & \mathcal{M}, w \vDash \neg\Diamond^{>}_{1}\phi \wedge \Diamond^{<}_{2}\phi \\
& \Leftrightarrow & \mathcal{M}, w \vDash \Diamond^{<}_{1}\phi \vee (\neg\Diamond^{>}_{1}\phi \wedge \Diamond^{<}_{2}\phi)
\end{array}$$

We show that $\mathcal{M}, w \vDash \Diamond^{<}_{1 \veebar 2}\phi \supset (\Diamond^{<}_{1}\phi \vee (\neg\Diamond^{>}_{1}\phi \wedge \Diamond^{<}_{2}\phi))$, and then that $\mathcal{M}, w \vDash (\Diamond^{<}_{1}\phi \vee (\neg\Diamond^{>}_{1}\phi \wedge \Diamond^{<}_{2}\phi)) \supset \Diamond^{<}_{1 \veebar 2}\phi$

☻

3.4.4 Expressing Conditionals

Notice also that the conditionals that we mentioned in §2.3.4 are definable here:

Remark 3.4.1:
If we introduce a next modality

$$\langle best \rangle$$

And define it explicitly with the following axiom, reducing it to the previous language

$$\langle best \rangle\psi \equiv k(\psi \wedge \neg\Diamond^{<}\top),$$

then we can easily obtain the following characterisation result (at the level of models, rather than just frames), where $\mathcal{F} = \langle \Omega, f^{<}, f^{>}, f^{k} \rangle$ is the 1-agent anonymous belief frame induced by any anonymous belief state $\preccurlyeq \subset W \times W$ and set Ω:

$$\langle \mathcal{F}, V \rangle \vDash \langle best \rangle\psi \Leftrightarrow min(W, \preccurlyeq) \cap [\![\psi]\!] \neq \varnothing,$$

More informally, $\langle best \rangle$ is true of a formula ψ just when there is a ψ-point which is minimal in the relevant pre-order. Now the semantics for the conditional $\phi \Box\!\!\Rightarrow \psi$ to which we alluded in §2.3.4 says roughly: at the best ϕ points, ψ holds. Its dual, $\phi \Diamond\!\!\Rightarrow \psi$ says that there is a best ϕ point at which ψ holds; i.e.:

$$\langle F, V \rangle \vDash \phi \Diamond\!\!\Rightarrow \psi \Leftrightarrow min([\![\phi]\!], \preccurlyeq) \cap [\![\psi]\!] \neq \varnothing$$

Remark 3.4.2:

$$\begin{aligned}
\mathcal{M} \models \phi \Diamond\!\!\Rightarrow \psi \quad &\Leftrightarrow \quad \mathcal{M}[!\phi] \models \langle best \rangle \psi \\
&\Leftrightarrow \quad \mathcal{M} \models [!\phi]\langle best \rangle \psi \\
&\Leftrightarrow \quad \mathcal{M} \models [!\phi]k(\psi \wedge \neg \Diamond^{<} \top) \\
&\Leftrightarrow \quad \mathcal{M} \models k(\phi \wedge [!\phi](\psi \wedge \neg \Diamond^{<} \top)) \\
&\Leftrightarrow \quad \mathcal{M} \models k(\phi \wedge [!\phi]\psi \wedge \neg [!\phi] \Diamond^{<} \top) \\
&\Leftrightarrow \quad \mathcal{M} \models k(\phi \wedge [!\phi]\psi \wedge \neg \Diamond^{<} (\phi \wedge [!\phi]\top)) \\
&\Leftrightarrow \quad \mathcal{M} \models k(\phi \wedge [!\phi]\psi \wedge \neg \Diamond^{<} (\phi \wedge \top)) \\
&\Leftrightarrow \quad \mathcal{M} \models k(\phi \wedge [!\phi]\psi \wedge \neg \Diamond^{<} \phi)
\end{aligned}$$

We do not know the structure of ψ, so cannot continue to use the PAL reduction axioms from §2.3.2, but for any given ψ there is a method to reduce a statement of the form $\phi \Diamond\!\!\Rightarrow \psi$ to the base language presented in this chapter.

We do not pursue further this logic. It is similar to a logic proposed in [8]. We have noted that one can express 'belief refinement' in such a logic.

3.4.5 Future Directions

One way to take future work here would be to flesh out the logic which was described briefly in this chapter. For example, it should be possible to modify it to talk about belief 'fusion' as described above, i.e. including *pedigreed* belief states. Remaining with the same language, one could also find other representation theorems for other natural merging operations of belief states. This should be straightforward, but would involve answering the second question with which we opened the last section.

That question (about different kinds of belief fusion or, more generally, merging) will be briefly raised again at the end of §4.3.3. A topic for future research is to determine what kind of mathematical structure would best capture the generality of different kinds of belief merging.

4 Rope Models

4.1 Introduction

Grove's representation theorem shows us that anonymous belief states are precisely the sort of structure that enable us to give a semantical characterisation of the AGM postulates. We now look at logical systems which allow for a multi-agent setting and whose semantics are based on such anonymous belief states.

The main reason for choosing an unorthodox presentation of an AGM representation theorem in §2.3.4 was to adopt a perspective which brings to the fore the similarities between sphere-system semantics and relational semantics. What are the points in a system of spheres model, and how should we enrich those models in order to make a serious claim towards an epistemic representation thesis? A point gives a propositional valuation, or a 'ground state', just as did points in relational models. The only other thing present is a well-behaved pre-order over the points, that represents the agent's disposition to revise her beliefs. Clearly more is needed. A system of spheres is a way to represent *one* belief state, where a belief state is taken to include dispositions to change the things actually believed (where the things 'actually believed' are given by the core of the belief state, $min(W, \preccurlyeq)$). In chapter 3 we have examined how such objects, i.e. systems of spheres not associated with particular points, can be interesting to study in a multi-agent setting; now we want to look at ways to represent several beliefs, including beliefs about beliefs.

In this chapter we will present two logical frameworks for reasoning about belief change. The first is Dynamic Doxastic Logic, or 'DDL'. DDL and its semantics are formulated in [31]. We devote some space to a discussion of repeated, or iterated, revision, add some axioms to DDL and showing their AGM postulate correspondents. We also continue the line of inquiry opened in §2.2.3 about the nature of the 'action' that the 'dynamic' operator represents in DDL; we present three different possible interpretations for the operator. Partly in connection with this inquiry, we attempt to follow up a line of thought of Krister Segerberg's in formulating a notion of 'complete belief', and we prove a completeness result for a logic including such an operator.

One of the three possible interpretations for the 'dynamic' operator in DDL might be called 'disposition to change' (rather than change *per se*). That is in effect the interpretation used in the second logical framework for reasoning about belief change, which we will call 'Doxastic Pre-encoding' Logic (DPL), and for whose formulation we rely largely on [7]. In that logic the 'dynamic' operator of DDL is re-cast as a 'static' dispositional operator: An operator for describing anonymous belief states. We then look at what more truly 'dynamic' operators are in this context, including for example the public announcement operator of PAL, and in each case we show whether and how the corresponding operator can be introduced into DDL.

4.2 Dynamic Doxastic Logic

Originally formulated as a modal logic to represent the AGM postulates, in their single non-introspective agent form, dynamic doxastic logic has been extended [31] to the introspective case, and here we present a version of it that we have extended in a natural way to the multi-agent case. The language of DDL has two types of modalities: doxastic and 'dynamic'. Thus for each agent a we will have a monadic operator B_a used to talk about the agents' beliefs, and for every formula ϕ with which it is possible to revise and for each agent a we will have a monadic operator $[*_a\phi]$.

4.2.1 Syntax and Semantics

The logic DDL is then given by the following axiom schemata and rules of inference:

Axioms:

(P) ϕ, where ϕ is a propositional tautology;

(K) $\Box(\phi \supset \psi) \supset (\Box\phi \supset \Box\psi)$, where $\Box \in \{B_a, [*_a\chi]\}$;

(R1) $[*_a\phi]B_a\phi$;

This corresponds to the AGM 'success' postulate, i.e. $(\dotplus 1)$ above; it says that after revising by ϕ, ϕ is believed. (See §4.2.2 below for the translation schema $AGM \leftrightsquigarrow DDL$.)

(R2) $[*_a\psi]B_a\bot \supset ([*_a\phi \wedge \neg\psi]B_a\bot \supset [*_a\phi]B_a\bot)$;

We have suppressed the K operator which is present in [31], expressing of its operand that the agent cannot revise by its negation. $K_a\phi \equiv [*_a\neg\phi]B\bot$ is a theorem of the logic of [31]; the K_a operators are thus completely definable. Here $[a_*\phi]B_a\bot$ can be read as $K_a\neg\phi$, but noted that it should *not* be read as 'agent a knows ϕ': We certainly do not have as an axiom $K_a\phi \supset \phi$, whereas such an axiom seems incontestable for a knowledge modality. [1] We give a K_a translation of this

[1] To put it in other terms, terms which are given a more formal meaning towards the end of this chapter: The use of K provides the reader with *soft* information that $K_a\phi$ should be read as saying that the agent knows (has hard information) that ϕ. However, as always with such information, the same

axiom in order to explain it: It is equivalent to $K_a\psi \supset (K_a(\psi \supset \phi)) \supset K_a\phi)$, i.e. to the K axiom for the compound modality $[*\neg -]B_\perp$.

(R3) $[*_a\neg(\phi \equiv \psi)]B_a\perp \supset ([*_a\phi]B_a\chi \equiv [*_a\psi]B_a\chi)$;

This axiom states that if to an agent there is no conceivable way in which two formulae ϕ and ψ could differ in truth value, then if that agent revises by those formulae it always has the same effect on the beliefs of that agent. It corresponds approximately to ($\dot{+}$4) above. We might now ask why the axiom does not have the form:

(R3') $[*_a\neg(\phi \equiv \psi)]B_a\perp \supset ([*_a\phi]\chi \equiv [*_a\psi]\chi)$.

After all, if an agent cannot conceive of two propositions differing in truth value, then why should they have different effects when she revises by them? For example, as it stands (R3) leaves open the possibility that $[*_a p \wedge p]\neg p$ and $[*_a p]p$, or $[*_a q]B_{a'}p$ and $[*_a q \wedge q]\neg B_{a'}p$. To decide which of the axioms (R3) or (R3') are appropriate, we should have a coherent interpretation of what the $[*\phi]$ operator is supposed to represent.

(R4) $[*_a\phi \wedge \psi]B_a\chi \supset [*_a\phi]B_a(\psi \supset \chi)$;

This is a translation of ($\dot{+}$5): It says that beliefs after revising by $\phi \wedge \psi$ include all of those beliefs that would be obtained by revising by ϕ and then adding (rather than revising by) ψ.

(R5) $[*_a\phi]b_a\psi \supset ([*_a\phi]B_a(\psi \supset \chi) \supset [*_a\phi \wedge \psi]B_a\chi)$;

This translates the postulate ($\dot{+}$6): If it after revising by ϕ, ψ is conceivably true, then we have the converse of (R4), i.e. that any belief obtained by revising by ϕ and then adding ψ will be a belief after revising by $\phi \wedge \psi$.

(R6) $[*_a\phi]\psi \equiv \neg[*_a\phi]\neg\psi$;

This expresses the functionality of the revision operator; it does not correspond to one of the AGM postulates, but is implicit in them.

(R7.1) $B_a\phi \supset [*_a\top]B_a\phi$

---

reader should be prepared to revoke beliefs based on it, for example when noting the awkwardness of the fact that as modelers we are not committed to saying that agent a is correct: $\neg\phi \wedge K_a\phi$ is consistent. All that $K_a\phi$ says for certain is that a cannot even conceive of being incorrect in a belief that ϕ. Of course, it would be possible to add some of

(RKT) $[*_a\phi]B_a\perp \supset \neg\phi$;

(RK4) $[*_a\phi]B_a\perp \supset [*_a[*_a\phi]b_a\top]B_a\perp$;

(RK5) $[*_a\phi]b_a\top \supset [*_a[*_a\phi]B_a\perp]B_a\perp$.

This would make the knowledge reading of $[*_a\neg\phi]B_\perp$ more acceptable. We do not pursue the topic here.

(R7.2) $b\top \supset ([*_a\top]B_a\phi \supset B_a\phi)$

These last two axioms concern revision by $\top$. The intuitive meaning of these axioms is as follows: The first (R7.1) states that beliefs cannot decrease when revising by a tautology: If you believe something before it, then you will believe that thing afterwards. The second (R7.2) states that if your beliefs are currently consistent then revising by a tautology will not increase your beliefs: That if after the revision you believe something, you believed it beforehand. We do not dwell here on what revision by $\top$ might mean, postponing such a question until we have established, as we attempt to in §4.2.3 what revision *tout court* might mean in DDL, and only raising it again for a technical reason in §4.2.4.

Rules of Inference:

$$\frac{\vdash \phi \qquad \vdash \phi \supset \psi}{\vdash \psi}\ MP;$$

$$\text{For } \Box \in \{B_a, [*_a\psi]\} : \frac{\vdash \phi}{\vdash \Box\phi}\ \Box N;$$

$$\frac{\vdash \phi}{\vdash [*_a\neg\phi]B_a\bot}\ KN;$$

$$\frac{\phi \equiv \psi}{[*\phi]\chi \equiv [*\psi]\chi}\ Sub$$

This last two rules are the only ones requiring any comment. As we remarked about the axiom (R2) above that it is the equivalent of axiom K for the compound modality $[*_a\neg-]B_a\bot$, so is KN the equivalent of rule N (necessitation) for that same modality. The final rule, substitution of provably equivalent formulae in revision contexts, is to avoid, for example, that $[*p]$ might have a different effect from $[*p \wedge p]$[2]. Notice that if we were to change (R3) to (R3') above, then this rule would be redundant, as it would be simulatable by KN, (R3') and MP.

We address now the semantics of the logic. The models of DDL are relational structures in which the belief- and revision- accessibility relations are related in an appropriate manner. That manner is appropriate in that it captures something of the AGM postulates: each point is associated with an anonymous belief state for each agent. It is formulated in terms of *selection functions* rather than pre-orders, but (as Lewis in effect shows[3]), the two definitions are effectively equivalent.

[2]This is why the logic is not suited for treating the dinosaur-vase scenario of Fermé and Hansson that we mentioned in §2.2.3.

[3][20] – On p. 49 he remarks the obvious equivalence, which we have taken for granted, between 'onions' and pre-orders; on p. 59 he proves the trickier equivalence between onions and selection functions. Note that unlike the sphere systems he considers, our onions are not 'centered'; [31] provides the proof found on p. 59 of [20] without the centering assumption.

Definition 4.2.1 (Selection Function):
Given a non-empty set W and a set $\mathbb{P}$ of propositions over W (i.e. subsets of W), F is a *selection function from* $\mathbb{P}$ *over* W iff$_{df}$:

(0) $F : \mathbb{P} \to \mathcal{P}W$;

(1) $F(X) \subseteq X$;

(2) $X \subseteq Y \Rightarrow (F(X) \neq \varnothing \Rightarrow F(Y) \neq \varnothing)$;

(3) $X \subseteq Y \Rightarrow (X \cap F(Y) \neq \varnothing \Rightarrow F(X) = X \cap F(Y))$.

(Cf. conditions (2)–(3) [20], p. 58.) The connection between selection functions and pre-orders is as follows: $min(P, \preccurlyeq_F) = F(P)$.

Definition 4.2.2 (DDL model):
Given a non-empty set W, an n-agent *DDL model* over W is a tuple

$$\langle W, \tau, \{f_a\}_{a \in n}, \{f_a^X\}_{X \in \tau}, V \rangle,$$

where

(a) $\langle W, \{f_a\}_{a \in n}, V \rangle$ is an n-agent relational model;

At each point, certain ground facts will hold, and each agent has certain beliefs, which, as in the case of the relational models (definition 2.3.3), are sets of points.

(b) τ is a Boolean algebra over W, whose elements we will refer to as τ-propositions. These should be thought of as the propositions by which agents are able to revise, or of which they can in some sense conceive: For example, one might let propositions in τ correspond to formulae ϕ for which there is a point w at which for some agent a, $w \models b\phi$.

(c) $f_a^X : W \to W$

Each agent also has a revision function for each of the elements of τ. This function effects a transition from one point to another. Thus revision is, like in *AGM*, completely deterministic. (Cf. Axiom (R6))

(d) For each $w \in W$, $f_a(f^-(w))$ is a selection function from τ over W.

That is the agent's new belief state after revision by ϕ is constrained in that it must be determined by some selection function. This is the locus of the relation between this semantics and the *AGM* postulates, or rather their semantic counterpart in the system of spheres, equivalent to selection functions. (Cf. [31])

(e) $f_a(w) \subseteq f_a(f^W(w))$

Semantical correspondent of axiom (R7.1).

(f) $f_a(w) \neq \varnothing \Rightarrow f_a(f^W(w)) \subseteq f_a(w)$
Correspondent of (R7.2).

We extend the truth definition for relational models to provide for the new $[*_a\phi]$ operators:

Definition 4.2.3 (Satisfaction of $\langle *_a\phi\rangle\psi$):
The $[*_a\phi]$ operators have the following satisfaction definition: $w \models [*_a\phi]\psi \Leftrightarrow_{df} f_a^X(w) \models \psi$, where X is the largest τ-proposition contained in $[\![\phi]\!]$.

I.e. we give the standard definition of satisfaction, given that the f^ϕ's are functional. So far we have presented the syntax and semantics of a multi-agent version of DDL. We will now extend existing work in a number of ways.

First of all we will show that it is possible to add axioms to the logic in order to characterise iterated revision. The AGM postulates say nothing about what the new anonymous belief state should look like after revising by a proposition: As we have seen, they only give the output belief set. Therefore they do not state how to go about revising iteratively. In §4.2.2 we give axioms that capture several ways to iterate revision based on the existing anonymous belief state.

In §4.2.3 we then go on to propose that the operation expressed by the modalities of DDL does not correspond sufficiently to the intuitive notion of an action: Because DDL is intended to be a modal logic for the AGM postulates, it, like them, does not consider revision as an action. Here we in effect recall §2.2.3, where we discussed for example introspection, and the problems generated by Moore sentences.

Before comparing DDL with what we call 'doxastic pre-encoding logic', we nonetheless provide, in §4.2.4 and then chapter 5, one technical addition to the existing DDL literature: A completeness proof for a 'complete belief' operator in DDL.

4.2.2 Iterated revision

Can we add axioms to DDL to ensure that iterated revision occurs in a certain way? Certainly we can, as long as that 'certain way' is expressible in the language. Then which 'certain way's are expressible in the language? We do not fully answer this question, but do provide an example of three natural ways, the first of which we have already seen, which are expressible in the language. Specifically, the first iteration strategy for which we will provide axioms is the *belief refinement* operation that we described in §3[4]. The

[4]In [7] it is the 'revolutionary' strategy; Segerberg has, in discussion, suggested that it might be called a 'trusting' strategy; cf. [27]. In the most comprehensive taxonomy of iterated belief revision strategies that I have seen ([25]), Rott calls this operator 'moderate revision'.

second we call, following [7], the 'Machiavellian' strategy[5] The last is the least rational, and we call it the 'cynical' strategy[6]. The axioms need to say things about the truth of formulae of the form

(It) $[*_a\phi][*_a\psi]\chi$.

Again, the interpretation of the revision operator is problematical here: Below we consider the special case that is formulae of the form $B_a\chi$. This will permit us to translate our axioms more directly into AGM postulates, using the translation schema used to generate the DDL axioms, i.e.:

$$[*_a\phi]B_a\psi \leftrightsquigarrow \Gamma \dotplus \phi \vdash \psi$$

For the iterated case, it looks like this:

$$[*_a\phi][*_a\psi]B_a\chi \leftrightsquigarrow (\Gamma \dotplus \phi) \dotplus \psi \vdash \chi$$

(We will use the right-to-left direction of these translation schemata in §4.2.2 to converting the Darwiche and Pearl postulates for iterated revision (§3.3.2) into DDL axioms.)

Refinement

Specifically, we want axioms to say for example that if it's coherent to a that ϕ and ψ both hold, then (It) holds just if after revising by the conjunction of ϕ and ψ, then χ holds. If, on the other hand, that is not coherent to a, then a simply forgets the original revision by ψ, and (It) is equivalent to $[*_a\phi]\chi$.

Then we have the axioms:

(DR1) $[*_a\phi \wedge \psi]b_a\top \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi \wedge \psi]B_a\chi)$;

(DR2) $[*_a\phi \wedge \psi]B_a\bot \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi]B_a\chi)$.

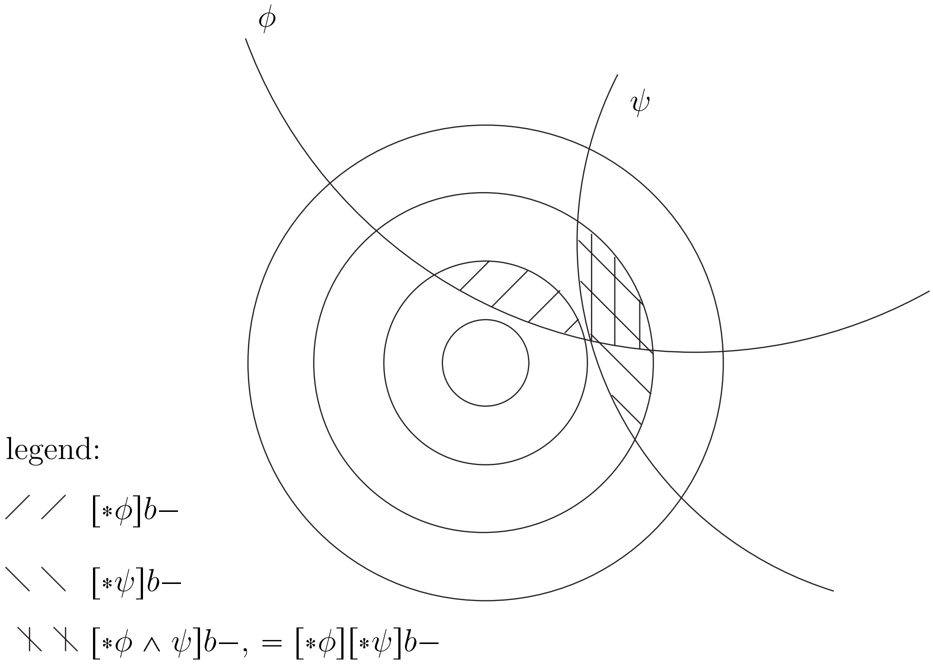
We might be tempted to replace $B_a\chi$ with χ:

(DR1') $[*_a\phi \wedge \psi]b_a\top \supset ([*_a\phi][*_a\psi]\chi \equiv [*_a\phi \wedge \psi]\chi)$;

[5] Segerberg's favoured term here is 'skeptical strategy'; again cf. [27]. Rott [25] calls it 'conservative revision' and 'natural revision'.

[6] This does not appear in Rott's typology; it is very similar to his 'restrained revision' though.

Figure 4.1: Refinement: The case when $[*_a\phi \wedge \psi]b_a\top$ holds:



(DR2') $[*_a\phi \wedge \psi]B_a\bot \supset ([*_a\phi][*_a\psi]\chi \equiv [*_a\phi]\chi)$.

Choosing (DR1−2') over (DR1−2) is difficult without a more concrete natural interpretation of the $[*_a\phi]$ operator. The same comment could apply to the other axioms we propose here for iterated revision.

We mentioned iterated revision in §2.2.4. Although we have argued that the *AGM* postulates appear to refer to what we call belief *sets* rather than to belief *states*, nonetheless notice that (DR1−2) could have been included as postulates:

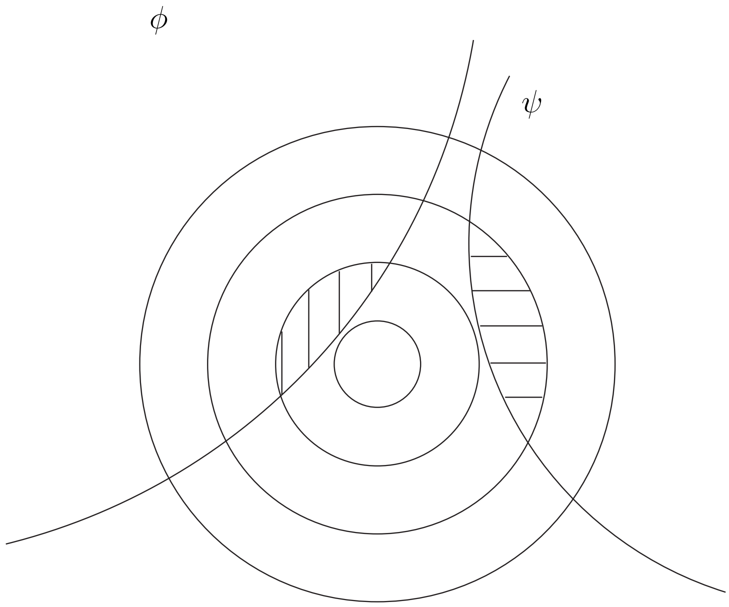
($\dotplus$DR1) $\Gamma \dotplus (\phi \wedge \psi) \nvdash \bot \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus (\phi \wedge \psi)$

($\dotplus$DR2) $\Gamma \dotplus (\phi \wedge \psi) \vdash \bot \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus \psi$

In this form it is possible to argue that they encapsulate a natural kind of minimality, or conservatism, of change.

Recall that we already had postulates, due to Darwiche and Pearle, which we proved were valid for this kind of iterated revision. Thus we have the following:

Figure 4.2: Refinement: The case when $[*_a\phi \wedge \psi]b_a\top$ does not hold:



Corollary 4.2.1:
(∔DR1-2), which characterise refinement[7], entail Darwiche and Pearl's (DP1-4) from §3.3.2.

and equally that the Darwiche and Pearl axioms, when translated into *DDL* will characterise the same iterated revisions:

Corollary 4.2.2:
The following

(LDP1) $[*_a\phi \wedge \neg\psi]B_a\bot \supset ([*_a\psi][*_a\phi]B_a\chi \equiv [*_a\phi]B_a\chi)$

(LDP2) $[*_a\phi \wedge \psi]B_a\bot \supset ([*_a\psi][*_a\phi]B_a\chi \equiv [*_a\phi]B_a\chi)$

(LDP3) $[*_a\phi]B_a\psi \supset [*_a\psi][*_a\phi]B_a\psi$

(LDP4) $[*_a\phi]b_a\psi \supset [*_a\psi][*_a\phi]b_a\psi$

are consequences of (DR1) and (DR2).

[7] Or 'Revolution', 'Trust', or 'Moderation' if you are van Benthem, Segerberg or Rott. 'Lexicography' if you have a less active imagination.

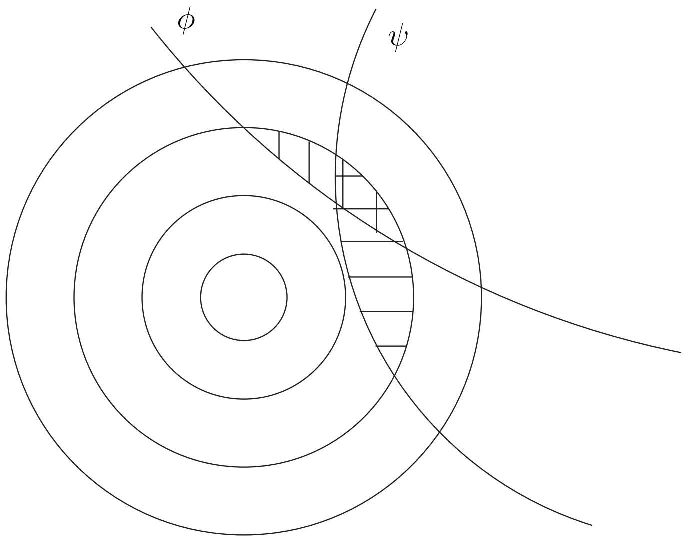
Machiavellianism

In the Machiavellian strategy, it is only the new belief set, obtained after revision, that is shifted and placed on the top; the others remain where they were, relative to each other. That is, it is only in the case where after revising by ϕ an expansion by ψ would be possible that the revision by ϕ before ψ affects that second revision:

(DM1) $[*_a\phi]b_a\psi \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi \wedge \psi]B_a\chi)$

(DM2) $\neg[*_a\phi]b_a\psi \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi]B_a\chi)$

Figure 4.3: Machiavellianism: The case when $[*_a\phi]b_a\psi$ holds:



In a sense that deserves to be made more precise, Machiavellianism is less 'stable' than Refinement.[8] That is, if two models of an agent's belief state are very *similar*, the Machiavellian strategy will not in general reflect that similarity. This instability could form the basis of an argument that the strategy is less rational.

Again, we can write postulates in the *AGM* format to express (DM1-2):

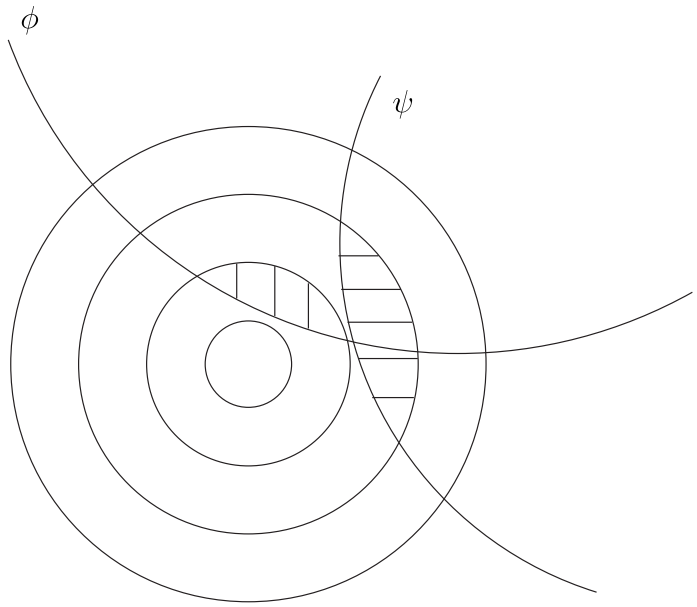
($\dotplus$DM1) $\Gamma \dotplus \phi \nvdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus (\phi \wedge \psi)$

($\dotplus$DM2) $\Gamma \dotplus \phi \vdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus \phi$

The final strategy that we present is also not 'stable', and certainly is less rational than both of the previous ones:

[8]It is Johan van Benthem who suggested this way to express my intuition here in terms of 'stability'.

Figure 4.4: Machiavellianism: A case when $[*_a\phi]b_a\psi$ does not hold:



Cynicism

The cynical strategy is in a sense like the Machiavellian strategy, in that it promotes the favoured ϕ points, after a revision by ϕ, to the top, but in the case where this belief state should turn out to be wrong, it quickly gives up and prefers to think that the original revision was incorrect. One form is strong cynicism:

(DSC1) $[*_a\phi]b_a\psi \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi \wedge \psi]B_a\chi)$

(DSC2) $[*_a\phi]B_a\neg\psi \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\psi \wedge \neg\phi]B_a\chi)$

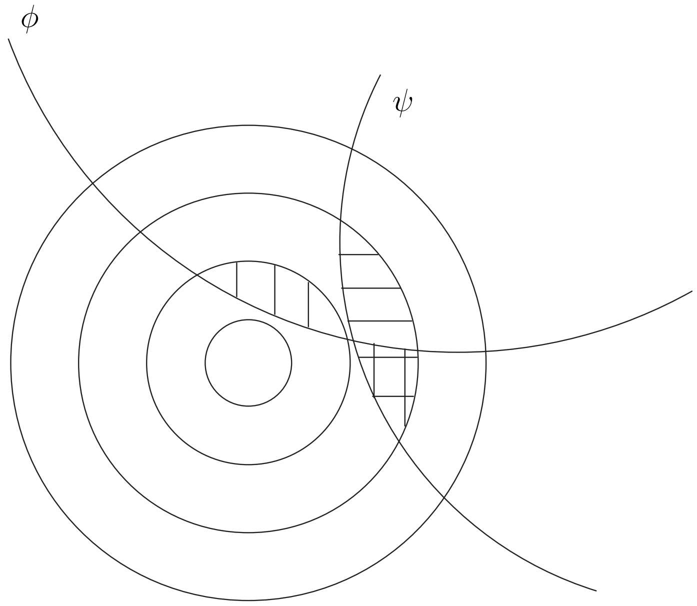
Here the agent will, after learning that a revision did not give her a correct theory, reject that revision in the strong sense that she will change her commitment set so as to rule out that the revision was correct.

In the weaker form, characterised by the following three axioms, after learning that a revision did not give her a correct theory, she will prefer to reject the revision over accepting it, but will not go so far as to modify her commitment set:

(DC1) $[*_a\phi]b_a\psi \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\phi \wedge \psi]B_a\chi)$

(DC2) $[*_a\phi]B_a\neg\psi \supset ([*_a\psi \wedge \neg\phi]b\top \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\psi \wedge \neg\phi]B_a\chi))$

Figure 4.5: The Cynical iterated revision strategy in action



$$\text{(DC3)}\quad [*_a\phi]B_a\neg\psi \supset ([*_a\psi \wedge \neg\phi]B\bot \supset ([*_a\phi][*_a\psi]B_a\chi \equiv [*_a\psi]B_a\chi))$$

Again, we can provide *AGM*-style postulates for this rather strange form of iterated revision:

$$(\dotplus\text{DC1})\quad \Gamma \dotplus \phi \nvdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus (\phi \wedge \psi)$$

$$(\dotplus\text{DC2})\quad \psi \nvdash \phi \Rightarrow (\Gamma \dotplus \phi \vdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus (\psi \wedge \neg\phi))$$

$$(\dotplus\text{DC3})\quad \psi \vdash \phi \Rightarrow (\Gamma \dotplus \phi \vdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi = \Gamma \dotplus \psi)$$

Discussion

Notice that the *DDL* axioms expressing these strategies all have a similar form. Indeed in (DR1) and (DM1) the consequent is the same; similarly for (DR2) and (DM2). This suggests, albeit in a very imprecise way, that there is some common structure to natural strategies, and one that might be exploited in addressing the research question raised at the end of §3 of finding general mathematical structures that express the full generality of belief *merging*, or indeed *fusion*.

Each of these sets of strategy-characterising axioms will restrict the class of frames to those in which the revision and belief operators are related in a specific way.

We do not here spell out more formally the correspondence between these axioms and their semantic correlates; some work has been completed here in [27], where similar axioms have independently been stated. We will look instead at the question we have skirted around a number of times so far, of just what the $[*-]$ operator is intended to represent DDL (in the sense in which the B_i and $[!G\phi]$ operators in PAL can be said, along the lines of Baltag and Moss' representation thesis, to represent factive belief and announcement to a group). This will lead us to look at another logic for revision, where such a formal correspondence is already established.

4.2.3 What is revision in DDL?

As Segerberg observes ([31], p. 89), the DDL way of representing revision does remain subject to the Moore problems we discussed in §2.2.3. I believe that there is some confusion about what is represented by the revision operator $[*_a\phi]$. Is it an action? Certainly the failure in the treatment of the Moore problems means that if it is an action it is not an announcement. There remain three clear possibilities for a coherent common-sense interpretation of the $[*_a\phi]$ operator.

1. It can be seen as an operator of the $STIT$ (See To It That) type, whereby the agent is forced to ensure that *after the action*, she believes ϕ. In this interpretation, which is most faithful to the intuitions of Segerberg, the action is in effect a command: "Make it the case that, or see to it that, you believe ϕ". In this case, clearly the Moore problem does not have the gravity which we have thus far assigned it. A command $[*_a p \wedge \neg B_a p]$ is simply a command which is impossible, for an agent with any normal (human-like) degree of introspection, to obey.

   An alternative way to phrase this interpretation is that the revision operator $[*_a\phi]$ does capture an action of announcement to a, by some source which is not (at least not *immediately*) doubted, but not of the proposition ϕ, but rather by the proposition 'It will be the case, *after this announcement*, that ϕ'.

   We describe below, in §4.2.4 a way in which the 'disbelief retention' problem can be eliminated, and in which the absurdity of the Moore sentence commands can be brought out within the logical system, even without adding anything like full introspection to any of the agents. (This would be of relevance to both this interpretation and the following one.)

2. If the operator is to be considered as some kind of announcement or statement, then the Moore sentences remain problematic. One option is to add some temporal element in the logic, specifically, a 'Yesterday' operator: $Y\phi$ means that, just before the last action, ϕ was true. Then the axioms could be altered so that anything of the form $[*_a\phi]\psi(B_a)$, where ψ is a formula containing an occurrence of the B_a operator, would be transformed into $[*_a\phi]\psi(B_aY)$, i.e. inserting a Yesterday

operator after each B_a operator, so that the first axiom, for example, would become

$$[*_a\phi]B_aY\phi$$

Which we would read, "After revising with the fact that ϕ, agent a believes that before that action ϕ held." This would require adding a temporal aspect to the semantics, and is only mentioned here as a possibility. We do address the possibility of adding time into the mix in chapter 7, but will not pursue further this interpretation.

3. Alternatively, the revision operator can be seen to be is a sort of *internal* action: A supposing, or an hypothesising, by the agent in question. Then $[*_a\phi]B_a\phi$ could for example be read as saying that if a were to suppose that ϕ, then a would be supposing a situation in which ϕ. Notice that the Moore problem does not appear here: I am perfectly able to entertain the possibility that it is raining outside and that I am unaware of this fact ($r \wedge \neg B_{me}r$); all that this means is that I believe that, *in that possibility which I am entertaining*, $r \wedge \neg B_{me}r$. Notice that the compositionality of the language is threatened by this interpretation: We do not say

   > When entertaining ϕ, then in the possibility which I am entertaining, I believe that ϕ holds,

   but rather

   > When entertaining ϕ, then I believe that, in the possibility which I am entertaining, ϕ holds.

   If we compare this with the previous suggestion for interpretation, involving a Yesterday operator, we can observe what may be merely a superficial similarity between the two. In each case, in order to make a compositional interpretation, of the standard recursively defined type, it would be necessary to insert an operator; this time the operator would express that the belief in question is not about the state of the world given the act of supposition which the agent has made, but rather about the object of the supposition. In the previous case, the operator would express that the belief in question is not about the state of the world given the act of announcement (or discovery, or whatever) that the agent underwent, but rather about the state before that act. This may be a superficial similarity, but nonetheless it should be clear that there is some similarity.

   Note that it is difficult to see here how to interpret nested revision modalities.

In all three of those interpretations, we can note now that it is unclear for example why $[*\phi]p$ is not an axiom of the system.

The language of DDL provides us with a succinct manner for expressing some of the richness of belief states. In §4.3, we will consider a more succinct but slightly less expressive language for representing belief states; there some of the subtleties of expression of DDL will be lost, but it will be precisely those parts which were problematic for the last two interpretations that we will no longer be able to express.

4.2.4 Complete Belief

We also note that it is possible to give an even richer (i.e. more expressive) language for representing belief states, by introducing an operator C that is a doxastic version of the 'common knowledge' operator familiar from the epistemic logic literature ([12]); in its multi-agent version we might call C a 'common belief' operator. Segerberg has suggested that this might provide us with a way to overcome the problems within DDL other than the Moore problem. That is, the problem of retained disbelief.[9] For complete belief we can give the following additional axioms:

(C1) $C_a\phi \equiv (B_a\phi \wedge B_aC_a\phi)$;

(C2) $C_a(\phi \supset B_a\phi) \supset (B_a\phi \supset C_a\phi)$;

We will present the semantics of this logic and prove completeness with respect to that. (The logic is of course not compact, so it is completeness in the sense of the logic capturing every validity of the form $\models \xi$ rather than $\Xi \models \xi$ where Ξ is an arbitrary set of formulae.) However, in order to do this we must first make some adjustments to the logic, which correspond to our worries about what the revision operator is supposed to represent. Specifically, we find that we are forced here to introduce the following axiom:

(R7!) $[*\top]\phi \equiv \phi$,

This in effect leads us to postulate the following:

($\dot{+}$7) $\Gamma \dotplus \top = \Gamma$,

which conflicts with ($\dot{+}$3), because it means that $\{\bot\} \dotplus \top = \bot$. To a critic of such an approach, one could reply in the following manner: The absurd belief state really is that, absurd. There is no way to recover from it. We might be able to imagine (or to recall) situations in which we were so surprised by a new piece of information, or command to believe something, that we were unable to know what to think afterwards. It might be thought that that is the absurd belief state. Then we could also imagine (recall) coming to regain beliefs, to find our epistemic feet. That would then be thought to be a contradiction with ($\dot{+}$7). However, this situation can be simulated in the logic DDL_C.

[9] Solving the Moore problems in this context would require dealing with the problem of giving a syntactic characterisation of successful updates (cf. [11]).

Specifically, one can introduce one or more propositional variables and give them an interpretation consistent with the informal description of the scenario just given. $p_\perp$, for example, might correspond in some way with 'I don't believe it!'.

We will also have to make a corresponding change in the frame class in order to make (R7!) sound.

We call the logic DDL with the three additional axiom schemata (C1), (C2) and (R7!) 'DDL_C'. Here we have given only single-agent versions of complete belief. To those familiar with the literature on common knowledge, this might seem an odd thing to be doing. In fact it is straightforward to extend the definitions to encompass groups of agents; the interest is rather in finding the transitive closure (and not the union of a collection) of belief relations.

Definition 4.2.4 (Transitive Closure (f^+)):
Given a function $f : W \to 2^W$, we denote the *transitive closure of* f as f^+, which is the smallest relation such that

1. $f(w) \subseteq f^+(w)$;
2. $w' \in f^+(w) \Rightarrow (f^+(w') \subseteq f^+(w))$.

This is non-trivial in this case because in DDL no significant assumptions are made about belief: it is not assumed to be positively or negatively introspective for example. Complete belief is the positively introspectible closure of belief. Clearly if the belief were already transitive at every point of the model (i.e. including after any revisions) then there will be no difference between the c and b operators.

Definition 4.2.5 (DDL_C model):
Given a non-empty set W, an *n-agent DDL_C model* over W is a tuple

$$\langle W, \tau, \{f_a^b, f_a^c, \{f_a^X\}_{X\in\tau}\}_{a\in n}, V\rangle,$$

where

- $\langle W, \tau, \{f_a^b, \{f_a^X\}_{X\in\tau}\}_{a\in n}, V\rangle$ is an n-agent DDL model, and

(d) $f_c = f_b^+$;

(e) $f_W = id$ – That is, the revision function induced by $\top$ is just identity: revision by $\top$ does nothing.

Theorem 4.2.1 (Completeness of DDL_C):
The logic DDL_C is complete for the class of DDL_C models.

Proof. Chapter 5

☻

4.3 Doxastic Pre-Encoding Logic

There is a more natural semantics than that of DDL for a language like that proposed at the end of §3.4. In this section we present that semantics, and describe its language and some of its logic; much of this resembles the formulation in [7]. We also make many comparisons between the logic and DDL.

4.3.1 Static Semantics

Definition 4.3.1 (n-agent Rope Model[10]):
An *n-agent rope model* is a tuple $\langle W, \{k_a\}_{a\in n}, \{\preccurlyeq_a\}_{a\in n}, V\rangle$, where $\langle W, \{k_a\}_{a\in n}, V\rangle$ is an n-agent relational model, and for each $i \in n$, $\preccurlyeq_a$: $W \rightarrow \mathcal{O}_{k_a(w)}$, (we write $\preccurlyeq_a^w$ for $\preccurlyeq_a (w)$).

In such a model, k_a plays the same role as in standard relational models: It gives what we can call the 'commitment set' of the agent (cf. [31]). It would be appropriate to write the modality corresponding to it as K_a, but in what follows we will not in fact make use of the commitment set. The new part in these kinds of models are the functions $\preccurlyeq_a$. At each point w and for each agent a, $\preccurlyeq_a^w$ represents a's belief state.

Note that putting $f_a(w) = min(W, \preccurlyeq_a)$ will let us write down, given an n-agent rope model, an n-agent relational model. Indeed, we put the following as the truth condition for the modal operator B_a in n-agent rope models:

(4') $w \models B_a\phi \Leftrightarrow min(W, \preccurlyeq_a^w) \subseteq [\![\phi]\!]$.

This means that the models we would write down would be indistinguishable, from the point of view of the language we are dealing with, from the original n-agent model. There is more information, of course, present in the n-agent rope model, and the question now (as it was in §3.4) is in what way we should enrich the logical language so that it can distinguish suitably between two models. The belief state of an agent at a point is given by the belief onion associated with that agent at that point. And what does such a belief onion give us? We can think of it as a series of 'fallbacks', representing what the agent would believe should she not believe some part(s) of her actual beliefs – represented by the core, $min(W, \preccurlyeq)$. In order to capture this information present in the model, we will

[10]We call them 'rope models' because 'rope' is the collective noun for (any rope of) onions; the 'Amsterdam canal model' suggested earlier would nonetheless have been more illustrative.

add a two-place modal operator to express belief, as is done in [7], which is semantically equivalent to Lewis' $\Box\!\!\Rightarrow$ operator. Where $B_a\phi$ 'said' that agent a believed ϕ, $\phi \boxed{a}\!\!\Rightarrow \psi$ 'says' roughly that if agent a believed ϕ, she would believe ψ. The formal definition is

(5) $w \models \phi \boxed{a}\!\!\Rightarrow \psi \Leftrightarrow min([\![\phi]\!], \preccurlyeq_a^w) \subseteq [\![\psi]\!]$.

There is enough information in a rope model to represent beliefs, including their disposition to change under the influence of various sorts of information.

The description given above of what $\phi \boxed{a}\!\!\Rightarrow \psi$ 'says' is thus 'rough' rather than precise because it does *not* mean that if the agent were to learn (for example, via an announcement) that ϕ then the agent would believe ψ. For example, it might be that from some point w, in the agent's nearest p-points, the agent does not believe that p (i.e. $w \not\models p \boxed{a}\!\!\Rightarrow \neg(\top \boxed{a}\!\!\Rightarrow p)$); yet after an announcement that p, the agent believes that p (i.e. $w \models [!p](\top \boxed{a}\!\!\Rightarrow p)$). We will examine some ways of formulating announcements, by the introduction of several different modal operators corresponding to expansions and revisions. Just before that, it is instructive to compare the static part of DPL with DDL.

4.3.2 Comparison with DDL

Notice that there is a similarity between the models for the static DPL which we have just seen, and the models for DDL. The difference is that what in a DDL model takes two modal steps in a rope model only takes one. That is: $[*_a\phi]B_a\psi$ can be read as saying: "Revise by ϕ, then go to all the belief states, then check that ψ holds", where $\phi \boxed{a}\!\!\Rightarrow \psi$ says, "Go to the closest ϕ points, then check that ψ holds". The DDL approach, while arguably a little unnatural[11], does *prima facie* allow more expressivity.

DDL allows formulae of the form

- $B_a\phi$,
- $[*_a\phi]p$,
- $[*_a\phi][*_{a'}\psi]\chi$,
- $[*_a\phi]B_aB_{a'}\psi$,
- $B_a[*_{a'}\phi]\psi$,

and so on, which are not in the general case directly translatable into DPL formulae.

[11]We see this unnaturalness reflected in the completeness proof of DDL_C that we provide (chapter 5), when we introduce a two-sorted model that reflects what is effectively a two-sorted approach.

We assume for the time being the 'suppositional action' interpretation of the DDL revision operator (described above in §4.2.3), which is, of the three interpretations we suggested, closest to the spirit of [7]. Then we will show that the extra expressivity of the language of DDL has no clear use under that interpretation.

Formulae of the form $B_a\phi$ can be expressed as $\top \boxed{a}\!\Rightarrow \phi$, and indeed this is how [7] expresses the basic monadic belief modality in terms of the binary one. In DDL there is the possibility for an agent to have incoherent belief states and to be 'resurrected' from this epistemic hell by revising by $\top$. I have no clear intuition about what this is supposed to mean. (Nonetheless we have made some comments about the absurd belief state above in §4.2.4.) If $[*\top]$ *were* to have such effects, this would raise similar issues as formulae of the form $[*_a\phi]p$, or more generall $[*_a\phi]\psi$, where ψ is a propositional formula. In such cases, there is certainly no intuition to be taken from [1] as to what should be the case, but there are good motivations for maintaining that after supposing it to be that case that (or even, in any of the more general senses, after revising by) ϕ, which might be taken to be an *epistemic* action, there is no change to such ground facts.

If we are happy to accept this account of how $B_a\phi$ can be expressed, then clearly we can express in the binary language formulae of the form $[*_a\phi]B_aB_a\psi$.

Consider now the case of formulae of the form $[*_a\phi][*_{a'}\psi]\chi$. Given that we are not interested in the case where χ is purely a Boolean combination of propositional variables, we are considering formulae of the form $[*_{a_0}\phi_0]\dots[*_{a_k}\phi_k]B_a\psi$. These are intended to represent iterated supposings. It is hard to see how my supposing, or entertaining, (presumably privately[12]) that ϕ could affect your belief state at all. Therefore we assume that $a_k = a$. For similar reasons, the only intelligible case, given this particular interpretation that we have chosen, is then where each of the $a_0 \dots a_k$ are a.

This in general represents iterated revision, which is discussed in §3.3.2 and 4.2.2. In the particular interpretation we have chosen to discuss here, it is iterated suppositions: If I entertain the idea that it is raining (r) and then entertain the idea that I am riding my bicycle (s), is this different from my entertaining the idea that it is raining and that I am riding my bicycle ($r \wedge s$)? If I entertain the idea that it is raining, and then entertain the idea that it is not raining, then I am certainly not entertaining the idea that is raining and not raining ($r \wedge \neg r$, i.e.$\bot$). In these cases of iterated entertainings, the three strategies discussed in 4.2.2 could be used to characterise different kinds of agents, or perhaps just different kinds of supposition.

The closest way of representing such a formula in the DPL language is $\phi_0 \boxed{a}\!\Rightarrow \dots \phi_k \boxed{a}\!\Rightarrow \psi$, i.e. $[*\phi_0]B_a[*\phi_1]\dots B_a[*\phi_k]B_a\psi$. Now if this were not the same as $[*\phi_0]\dots[*\phi_k]B_a\psi$ then it would be because our model represents an epistemic state in which the agent is not introspective about her own belief state, where 'belief state' now is taken to include dispositions to change beliefs. This is certainly an interesting topic, and is related to the

[12] Perhaps there is some scope for developing a 'Public Entertainment Logic', but this is something we will not pursue here

issue of 'complete belief', for which we provide a completeness proof in DDL. However, we will not say much more about it here. We will return briefly to the issue during our presentation of the dynamic part of DPL. (We do not address at all the question what formulae of the form $[*_{a_0}\phi_0]\ldots[*_{a_k}\phi_k]B_a\psi$ should be taken to mean when the a_i's are distinct.)

Consider now the last example of formulae not expressible in the language of DPL but expressible in that of DDL: $B_a[*_{a'}\phi]\psi$. Again, we are not interested in the case where ψ is of the form $B_{a'}\chi$. Then once again it is very unclear, *if* $[*_{a'}\phi]$ is to be interpreted as a private epistemic action by a', just what such a formula should express. For example, the formulae $B_a[*_{a'}\phi]p$ 'says' that a believes that if a' were to entertain ϕ, then p would hold. $B_a[*_{a'}\phi]B_a p$ is similarly difficult to attribute with sense.

DDL is undeniably a more expressive language than that of DPL, and for talking about models which are very similar to rope models, given the equivalence between pre-orders and selection functions. In the semantics of DDL it is the selection functions which are the central representations of the doxastic states of the agents. Those selection functions are, as we saw in the semantics, made out of a compound of two functions each corresponding to one modal operator. We saw that a discussion of iterated revision was possible in the DDL case. We will now look at how a similar approach is possible in DPL.

4.3.3 True Dynamics

In [7], van Benthem considers two kinds of belief change: under 'hard' and 'soft' information. The epistemic change in DEL can be considered to be 'hard' belief change, and is comparable to Segerberg's 'irrevocable' belief revision: Parts of the model are deleted, just as in PAL. This can be seen as belief expansion. The epistemic change under soft information is closer to belief revision of the AGM tradition: The ordering between points is changed. Each of them is expressed in the language by a modality of the same type as in PAL and DDL: i.e. a monadic modality for each proposition. The new language is 'reduced' to the static one by means of PAL-style reduction axioms. There is a great deal of expressivity in the static language, to the point that very specific forms of revision, including the one defined by refinement (§3.3.2) can be defined in the language.

Hard Change – public announcement DPL

The effect of a public announcement in the DPL semantics is exactly the same as it was in the relational semantics context of §2.3.2. That is, an announcement of a fact ϕ eliminates from the model all points at which ϕ does not hold. The additional reduction axiom looks like this (cf. [7], p. 12):

(PR) $[!\phi]\psi \boxed{a}\!\!\Rightarrow \chi \equiv \phi \supset ((\phi \wedge [!\phi]\psi) \boxed{a}\!\!\Rightarrow [!\phi]\chi)$

[7] shows the soundness of that axiom. We can give such reduction axioms for the DDL operators of which DPL's binary operator is a compound. However we will see that they do not have the same compound effect as that one reduction axiom for DPL. Consider the following two reduction axioms for the 'static'[13] language of DDL:

1. $[!\phi]b_a\psi \equiv \phi \wedge b_a(\phi \wedge [!\phi]\psi)$
2. $[!\phi][*_a\psi]\chi \equiv \phi \wedge [*_a\phi \wedge [!\phi]\psi][!\phi]\chi$

We can similarly show the soundness of each of these parts. (1) is basically (GG4), the reduction axiom for public announcements provided in 2.3.2. The soundness of the second can by seen by noting that

$$\begin{array}{rcl} w \vDash_{\mathfrak{M}} [!\phi][*_a\psi]\chi & \Leftrightarrow & w \vDash_{\mathfrak{M}[!\phi]} [*_a\psi]\chi \\ & \Leftrightarrow & w \vDash_{\mathfrak{M}} \vDash \phi \ \& \ f^{[\![\psi]\!]^{\mathfrak{M}[!\phi]}} w \vDash_{\mathfrak{M}[!\phi]} \chi \\ & \Leftrightarrow & w \vDash_{\mathfrak{M}} \vDash \phi \ \& \ f^{[\![\psi]\!]^{\mathfrak{M}[!\phi]}} w \vDash_{\mathfrak{M}} [!\phi]\chi \\ & \Leftrightarrow & w \vDash_{\mathfrak{M}} \vDash \phi \ \& \ f^{[\![\phi \wedge [!\phi]\psi]\!]^{\mathfrak{M}}} w \vDash_{\mathfrak{M}} [!\phi]\chi \\ & \Leftrightarrow & w \vDash_{\mathfrak{M}} \vDash \phi \ \& \ w \vDash_{\mathfrak{M}} [*\phi \wedge [!\phi]\psi][!\phi]\chi \\ & \Leftrightarrow & w \vDash_{\mathfrak{M}} \vDash \phi \wedge [*\phi \wedge [!\phi]\psi][!\phi]\chi. \end{array}$$

Putting (1) and (2) together gives us

$$[!\phi][*_a\psi]b_a\chi \equiv \phi \wedge [*_a(\phi \wedge [!\phi]\psi)](\phi \wedge b_a(\phi \wedge [!\phi]\chi)).$$

Notice that we cannot then proceed to obtain the axiom (PR), or any minor modification of it. This is because of the strange, almost two-sorted nature of the DDL models[14]. This two-sortedness, which we have already mentioned in §4.3.2, is exemplified in the following paragraphs.

Certain worlds in a model might only be used for revision, and not for beliefs. We give a formal definition to make this clear:

Definition 4.3.2 (Revision world):
A world w in a DDL model is a *revision world* iff$_{df}$ there is no world w' in that model and agent a such that $w \in f_a(w')$.

[13] There is a problem of nomenclature here: The word 'dynamic', sounds very fun, but its meaning is confused, as it is being used in different ways. In van Benthem's usage, 'dynamic' appears to mean 'model-changing', whereas whereas this is not the case in Segerberg's 'dynamic doxastic logic'.

[14] This two-sortedness is exemplified when we exploit it in the completeness proof in the next chapter.

Take a model in which a point r is a revision world. It is clear that for any proposition letter p, any formulae ϕ and ψ, and for any world w, in order to determine whether $w \vDash [*_a\phi]B_a\psi$ it is irrelevant whether or not $r \in V(p)$:

Proposition 4.3.1 (Irrelevance of valuations for revision worlds):
Take any model $\mathfrak{M} = \langle W, R, V \rangle$ containing a revision world $r \in W$. Then define the model $\mathfrak{M}' = \langle W, R, V' \rangle$ where for any $p \in \Phi$ and any $s \neq r$, $s \in V'(p) \Leftrightarrow s \in V(p)$, and $r \in V'(p)$ decided arbitrarily by tossing a coin. The following are equivalent for any agent a, formula ψ and $w \in W$:

1. $w \vDash_{\mathfrak{M}} [*_a\phi]B_a\psi$
2. $w \vDash_{\mathfrak{M}'} [*_a\phi]B_a\psi$.

Then we can have two such models $\mathfrak{M}$ and $\mathfrak{M}'$, who only differ in their evaluation of the revision world r. Suppose for example that $r \in V(p)$ but $r \notin V'(p)$. Then $r \in \mathfrak{M}[!p]$ but $r \notin \mathfrak{M}'[!p]$. Entailing that there *can* be some formula ψ and some world w such that

1. $w \vDash_{\mathfrak{M}[!p]} [*_a\phi]B_a\psi$ but
2. $w \nvDash_{\mathfrak{M}'[!p]} [*_a\phi]B_a\psi$.

Then there is clearly no way to reduce, via a single reduction axiom, the compound $[!\phi]([*_a\phi]B_a\psi)$. Two reduction steps are needed. Again, it is unclear to me what could be the advantage of such a complication that must be present in such a dynamic[15] version of DDL.

Soft Change – public revision DPL

In hard change, points are irrevocably eliminated from the model, representing the acquisition of absolute ('hard') information to the effect that a certain proposition holds. There are other more subtle forms of information change that we might consider, which alter the agents' orderings over points without irrevocably eliminating them. [7] gives reduction axioms for two very specific changes in orderings. One of them, there called 'Revolutionary', or 'lexicographic upgrade', is again the *belief refinement* of §3.

There are reduction axioms to the existing language that enable us to characterise these very specific relation changes. For example, consider a modal operator for the *action*[16] of belief refinement (lexicographic upgrade), $[\Uparrow\phi]$, true of a formula ψ just when after

[15] Here we are using 'dynamic' in Baltag's sense: See §6.3.4 below.

[16] This is not the same as the characterisation provided in §4.2.2, for reasons (which I hope are by now at least not entirely foreign) relating to the static nature of the logic in §4.2.2. Of course, there is a big similarity between the actual axioms.

refinement by ϕ, ψ holds. The (only interesting) reduction axiom, that for the binary belief modality, is as follows (cf. [7]):

$$\text{(RR)}\ \ [\Uparrow\phi]\psi \boxed{a}\!\!\Rightarrow \chi \equiv \left(\begin{array}{c} (\neg((\phi \wedge [\Uparrow\phi]\psi) \boxed{a}\!\!\Rightarrow \bot) \wedge (\phi \wedge [\Uparrow\phi]\psi) \boxed{a}\!\!\Rightarrow [\Uparrow\phi]\chi) \\ \vee \\ ((\phi \wedge [\Uparrow\phi]\psi) \boxed{a}\!\!\Rightarrow \bot \wedge [\Uparrow\phi]\psi \boxed{a}\!\!\Rightarrow [\Uparrow\phi]\chi) \end{array} \right)$$

Other such ordering-changing operations are possible, and many will be definable in the language as it stands. However, there is something a little *ad hoc* about these reduction axioms, and it would be of great interest to find a general logic in which such ordering changes could be expressed. Just as DEL generalises the relation changes of PAL, the language of [4] allowing such relation changes (which represent actions) to be expressed in the language in a completely axiomatised manner, it should be possible to do the same here with ordering changes.

This is not something that we can pursue here however, leaving it open as a topic for future research.

In this (non-eliminative) case it *is* possible to give reduction axioms of the $[\Uparrow A]$ operator to the language of DDL. One might hope that the DDL approach with its static revision operator and standard monadic belief operator would allows a clearer exposition of the reduction axioms for specific order-changing functions like (RR). However, as we will see there is not much additional clarity to be had. The relevant reduction axioms of the $[\Uparrow\phi]$ operation for DDL are as follows:

$$\text{(RR1)}\ \ [\Uparrow\phi][*_a\psi]\chi \equiv \left(\begin{array}{c} ([*_a\phi \wedge [\Uparrow\phi]\psi]b_a\top \wedge [*_a\phi \wedge [\Uparrow\phi]\psi][\Uparrow\phi]\chi) \\ \vee \\ ([*_a\phi \wedge [\Uparrow\phi]\psi]B_a\bot \wedge [*_a[\Uparrow\phi]\psi][\Uparrow\phi]\chi) \end{array} \right)$$

What does (RR1) say? If we brush aside some of the recursive details, looking at an instance where ϕ and ψ are both proposition letters, we get:

$$\text{(RR1')}\ \ [\Uparrow A][*_a p]\chi \equiv \left(\begin{array}{c} ([*_a A \wedge p]b_a\top \wedge [*_a A \wedge p][\Uparrow A]\chi) \\ \vee \\ ([*_a A \wedge p]B_a\bot \wedge [*_a p][\Uparrow A]\chi) \end{array} \right)$$

As in the case of (RR) we have a disjunction, two cases: If at w the agent can conceive of some world in which A and p hold, then $f^{[\![p]\!]}(w)$ will be $f^{[\![p\wedge A]\!]}(w)$; if not, it will remain as it is.

Now (RR2) can be obtained bearing in mind axiom (R7). Then

$$\text{(RR2)}\quad [\Uparrow\phi]B_a\chi \equiv \begin{pmatrix} ([*_a\phi]b_a\top \wedge [*_a\phi]B_a[\Uparrow\phi]\chi) \\ \vee \\ ([*_a\phi]B_a\bot \wedge B_a[\Uparrow\phi]\chi) \end{pmatrix}$$

This axiom again has the disjunctive form: If agent a can conceive ψ possible, then she will believe χ after upgrading all the ϕ worlds iff in the closest ϕ worlds she believes (the proposition expressed in the new model by) χ. If not, then her belief remains unchanged (modulo the last parenthetical recursive caveat).

In the next chapter we provide a technical intermezzo: The completeness for DDL_C. Skipping to chapter 6, we turn our attention to an algebraic approach to representing belief change. After a brief presentation of that approach we will apply some of the concepts from this chapter to illuminate the problem of trying to axiomatise the algebraic revision function.

5 Completeness of DDL_C

5.1 Introduction

We will show that the logic DDL_C is complete with respect to the class of its models, i.e. that given a consistent formula ξ, there is a DDL_C model containing a point that satisfies ξ. We do this for the single-agent case because the notation becomes heavy otherwise. The result can be extended in a straightforward way to the many-agent case.

For the slightly simpler notation which is possible in the single-agent case, we redefine DDL_C models (cf. definition 4.2.5).

Definition 5.1.1 (DDL_C model):
Given a non-empty set W, a DDL_C *model* over W is a tuple

$$\langle W, \tau, f_b, f_c, \{f_X\}_{X \in \tau}, V \rangle,$$

where

- $\langle W, \tau, f_b, \{f_X\}_{X \in \tau}, V \rangle$ is a 1-agent DDL model, and

(g) $f_c = {f_b}^+$;

5.2 Building a Model

We will define several different models, each one building on features of the last. The fourth model we arrive at will be the one we want: it will contain a point that makes ξ true, and furthermore it will be a model of the requisite kind.

5.2.1 Parsonical model

The first model we define will be isomorphic to the smallest filtration of the full canonical model for DDL_C. We define it, as in [9], in terms of 'atoms', that is: maximally consistent subsets of a (sub-formula-closed) set of formulae. That set of formulae will be 'Π-closed'. That is, sub-formula-closed in the following sense:

Definition 5.2.1 (Π-closure):
Given a set Σ of formulae, its Π-closure, $\Pi(\Sigma)$, is the smallest set such that:

1. $\Sigma \subseteq \Pi(\Sigma)$
2. $\neg\phi \in \Pi(\Sigma) \Rightarrow \phi \in \Pi(\Sigma)$
3. $\phi \bigcirc \psi \in \Pi(\Sigma) \Rightarrow (\phi \in \Pi(\Sigma)\ \&\ \psi \in \Pi(\Sigma))$, where $\bigcirc \in \{\vee, \wedge, \supset\}$
4. $\Diamond\phi \in \Pi(\Sigma) \Rightarrow \phi \in \Pi(\Sigma)$, where $\Diamond \in \{b, c, [*\phi]\}$
5. $[*\phi]\psi \in \Pi(\Sigma) \Rightarrow \phi \in \Pi(\Sigma)$
6. $(\phi \in \Pi(\Sigma)\ \&\ \forall\psi \in \mathcal{L}, \phi \neq \neg\psi) \Rightarrow \neg\phi \in \Pi(\Sigma)$

In what follows we assume Υ to be the Π-closure of some set Σ of formulae.

Remark 5.2.1:
When Σ is finite, $\Pi(\Sigma)$ is finite.

We write M_L for the set of maximally DDL_C-consistent sets of sentences of the language of DDL_C. We write $M_{L/\Upsilon}$ for the set of equivalence classes of M_L over Υ. We write $\mathcal{A}_\Upsilon$ for the set of maximally consistent subsets of Υ.

Remark 5.2.2:
There is a natural isomorphism $\alpha : \mathcal{A}_\Upsilon \to M_{L/\Upsilon}$: the connection is $\alpha(\Gamma) = \{\Xi \in M_L | \Gamma \subseteq \Xi\}$.

This remark not essential to the proof that follows, but the reader might find it useful, when reading the following proof, to think of the members of $\mathcal{A}_\Upsilon$ at once as sets of formulae (finite sets when Υ is finite; and henceforth we assume that Υ is finite) and as (infinite) subsets of M_L, the domain of the familiar canonical model.

Sets of sets of formulae are thought of as propositions, and, for any set of sets of formulae $\mathcal{A}$, we define the function:

Definition 5.2.2 ($/-/$ (propositions over $\mathcal{A}$)):

$$\begin{array}{lccl} /-/ & \Upsilon & \to & 2^{\mathcal{A}} \\ & \phi & \mapsto & \{\Gamma \in \mathcal{A} | \phi \in \Gamma\} \end{array}$$

(It will never be ambiguous which set $\mathcal{A}$ is question, thus we have suppressed the more rigorous notation $/-/_{\mathcal{A}}$.)

We will have to define the algebra of propositions τ over the domain of the model. It will be useful in the proof that follows to be sure that each member of τ is some $/\phi/$, where for every $\Gamma \in \mathcal{A}_\Upsilon$, $\Gamma \vdash \phi \Leftrightarrow \Gamma \nvdash \neg\phi$. Therefore we will make use of the following definition, which we will use to make τ be the set of Boolean combinations of Υ-propositions over $\mathcal{A}$:

Definition 5.2.3 ($-^{\curlyvee}$):
$\Upsilon^{\curlyvee} =_{df}$ the smallest set such that

1. $\Upsilon \subseteq \Upsilon^{\curlyvee}$;
2. $\phi \in \Upsilon^{\curlyvee} \Rightarrow \neg\phi \in \Upsilon^{\curlyvee}$;
3. $\{\phi, \psi\} \subseteq \Upsilon^{\curlyvee} \Rightarrow \phi \bigcirc \psi \in \Upsilon^{\curlyvee}$, where $\bigcirc \in \{\vee, \wedge, \supset\}$.

Lemma 5.2.1:
For any $\phi \in \Upsilon^{\curlyvee}$ and any $\Gamma \in \mathcal{A}_\Upsilon$, $\Gamma \vdash \phi \Leftrightarrow \Gamma \nvdash \neg\phi$.

Proof. By induction on ϕ. The base case is immediate. The inductive hypothesis is that for any formula ψ smaller than ϕ, $\Gamma \vdash \psi \Leftrightarrow \Gamma \nvdash \neg\psi$. Then the case where $\phi = \neg\psi$ is immediate; we show just

$\wedge$: $\phi = \psi \wedge \chi$.

- $\Rightarrow$: Immediate from consistency of Γ;
- $\Leftarrow$: By contraposition: Assume that $\Gamma \nvdash (\phi \wedge \psi)$. Then $\Gamma \nvdash \phi$ or $\Gamma \nvdash \psi$. So by inductive hypothesis, $\Gamma \vdash \neg\phi$ or $\Gamma \vdash \neg\psi$; either way, $\Gamma \vdash \neg(\phi \wedge \psi)$.

☻

That detail aside, we can now define the first model, before seeing why it is not adequate for our purpose.

Definition 5.2.4 (Parsonical Model over Υ):
The *parsonical model* over Υ is the tuple $\mathbb{A}_\Upsilon$:

$$\langle \mathcal{A}_\Upsilon, \tau, \{g_\diamond\}_{\diamond \in \{b,c,[*\phi] | \phi \in \Upsilon\}}, V \rangle,$$

where:

- for $\diamond \in \{b, c\}$, $g_\diamond(\Gamma) = \{\Delta | \nvdash \bigwedge \Gamma \supset \neg b \bigwedge \Delta\}$;
- $\tau = \{/\phi/ | \phi \in \Upsilon^{\curlyvee}\}$; ; and
- $V(p) = \{\Gamma \in \mathcal{A}_\Upsilon | p \in \Gamma\}$.

Ignoring for a parenthetical moment the algebra of propositions τ, the model is, as we have already noted, isomorphic to the smallest filtration of the full canonical model. Using such a filtration is the technique applied in Segerberg's completeness proof of PDL [29], to overcome a similar problem of non-compactness. A very similar approach to the one we have taken so far is used in [19] and in the textbook presentation of [9], to tackle the same problem.

However, the reason that our completeness proof is not (more) immediate is that information from the canonical model is needed regarding the behaviour of the revision functions in order to obtain a model that respects AGM, i.e. that respects the condition that each $f_b(f_-(w))$ be a selection function (cf. definition 4.2.2).

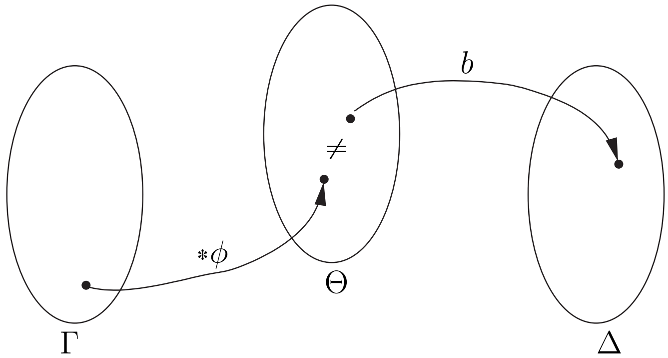


Figure 5.1: An example of why the Parsonical model is inadequate.

For example, situations like the one depicted in figure 5.1 are possible, indeed commonplace, in the parsonical model, and this prevents the AGM conditions from being respected there.[1] – In that illustration for example, we do not necessarily have $\Delta \vdash \phi$, as the first selection function clause would require. Yet we want to use a finite model in order to exploit one of the known techniques for obtaining a model in which the common belief function behaves correctly with respect to the belief function. In the parsonical model, we do not even have this. Yet it is obtainable, because such finite subsets are definable in terms of a single formula[2]. So our strategy will be to build a 'two-sorted'

[1] In fact, things are worse than that: The situation depicted in figure 5.1 can be remedied by using the closure condition

$$\frac{[*\phi]\psi}{b\phi \quad b\neg\phi}$$

in Π, including also $[*\top]\top$, and then defining τ as $\{/\phi/ | \phi \in \{\psi | b\psi \in \Upsilon\}^{\smile}\}$. Such an approach looked promising once on a train to Wassenaar. However, the final condition of selection functions remains intractable with that approach.

[2] See lemma 5.2.2; this technique is discussed in [28], pp.31f. We could note that it enables us to use the following form of reasoning in our proofs:

$$\frac{\Gamma \in f_{\Diamond}(w)}{w \vdash \Diamond\Upsilon} .$$

model, in which we have points corresponding to the points in the parsonical model, used to obtain a model in which the common belief function behaves correctly, and points associated with maximally consistent sets of formulae (points in the canonical model), which we will use to show that our model respects AGM.

5.2.2 Orthodox Model

We will now enrich the language, but in a conservative manner: We introduce a fresh propositional variable[3], p_∞. We do this in order to distinguish between the members of the two sets that will make up the domain of our model. p_∞ will be a member of precisely those points which correspond to full maximally consistent sets of formulae.

One half of the domain of the orthodox model, then, is $\mathcal{B}_\Upsilon = \{\Delta \cup \{\neg p_\infty\} \mid \Delta \in \mathcal{A}_\Upsilon\}$.

The other half takes a little bit more defining. We need to have representatives for all of the points which are ϕ-accessible for some ϕ. The vagueness of this definition hints that we need a recursive solution: Given $\mathcal{B}_\Upsilon$, we first define the points that are ϕ-accessible, then take these points and define those which are ϕ-accessible from there, and so on. Now we remarked already that each of these points will be represented by a point in the canonical model, so that there is an obvious way to do the inductive step described. There is a problem, however, in getting the ball rolling: How do we pick *one* possible outgoing ϕ arrow from some Γ, when several are possible? It is not enough to pick just one: We must pick one that is somehow related to any other outgoing ψ arrow that we pick. The solution to this little part of the puzzle, taken from something Michael Strevens achieved in a slightly different setting (see [30], pp.166-167), is just to define a choice function, that picks a member of the canonical model for each member of the parsonical model.

Definition 5.2.5 (Parson's Choice function):

Any function $\begin{array}{lllll} \sigma: & \mathcal{A}_\Upsilon & \rightarrow & M_L \\ & \Gamma & \mapsto & \sigma(\Gamma) \supset \Gamma \end{array}$ is a *Parson's Choice* function.

The Parson's Choice will in effect pick from the canonical model the revision dogma for the orthodox model to follow. We can actually use any such function, so pick one, σ.

Now to move from one step to the next in the inductive definition that is to follow, we need a function corresponding the the ψ-revision functions from the canonical model:

Definition 5.2.6 (h_ϕ):

[3] Or we could just take one from the language which has not been used in the (finite) formula ξ.

$$\begin{array}{llll} h_\phi: & M_L & \to & M_L \\ & \Gamma & \mapsto & \{\delta | [*\phi]\delta \in \Gamma\} \end{array}$$

Here we in effect take the restriction of the appropriate functions from the canonical model. It is a sound definition because the full canonical model is functional.

Now we can define the other part of the domain of the orthodox model, which we'll call $\mathcal{Y}_\Upsilon$. We define it recursively, the initial step requiring that we have already defined $\mathcal{A}_\Upsilon$:

$$\begin{array}{lll} \mathcal{Y}_{\Upsilon,0} & = & \{h_\phi(\sigma(\Gamma)) \cup \{p_\infty\} | \phi \in \Upsilon, \Gamma \in \mathcal{A}_\Upsilon\} \\ \mathcal{Y}_{\Upsilon,i+1} & = & \{h_\phi(u - \{\neg p_\infty\}) \cup \{p_\infty\} | \phi \in \Upsilon, u \in \mathcal{Y}_{\Upsilon,i}\} \\ \mathcal{Y}_\Upsilon & = & \bigcup_{i \in \mathbb{N}} \mathcal{Y}_{\Upsilon,i} \end{array}$$

(Note that this $\mathcal{Y}_\Upsilon$ is often infinite even though Υ is finite; this is not important for our concerns.)

The domain of the orthodox model will be $\mathcal{U}_\Upsilon = \mathcal{B}_\Upsilon \cup \mathcal{Y}_\Upsilon$. Extend σ to be the following function:

Definition 5.2.7 (Orthodox Choice):

$$\begin{array}{llll} \varsigma: & \mathcal{U}_\Upsilon & \to & M_L \\ & \Gamma \in \mathcal{B}_\Upsilon & \mapsto & \sigma(\Gamma - \{\neg p_\infty\}) \\ & u \in \mathcal{Y}_\Upsilon & \mapsto & u - \{p_\infty\} \end{array}$$

Remark 5.2.3:
For any $\phi \in \Upsilon$ and any $u \in \mathcal{U}_\Upsilon$, $\phi \in \varsigma(u) \Leftrightarrow \phi \in u$, i.e. $\varsigma(u)$ is an Υ-extension of u.

We are ready to define the next model:

Definition 5.2.8 (Orthodox Model over Υ):
The *orthodox model* over Υ is the tuple $\mathbb{S}_\Upsilon$

$$\langle \mathcal{U}_\Upsilon, \tau, \{j_\diamond\}_{\diamond \in \{b,c\}}, \{f_X\}_{X \in \tau}, V \rangle,$$

where:

1. $\mathcal{U}_\Upsilon = \mathcal{B}_\Upsilon \cup \mathcal{Y}_\Upsilon$.
2. $\tau = \{/\phi/ | \phi \in \Upsilon^{\curlyvee}\}$.
3. For $\Diamond \in \{b, c\}$:
    a) For $\Gamma \in \mathcal{B}_\Upsilon$, $j_\Diamond(\Gamma) = \{\Delta \in \mathcal{B}_\Upsilon | \Gamma \nvdash \neg b \bigwedge \Delta\}$;

b) For $u \in \mathcal{Y}_\Upsilon$, $j_\Diamond(u) = \{\Delta \in \mathcal{B}_\Upsilon | u \vdash b \bigwedge(\Delta - \{\neg p_\infty\})\}$

4. • For any $\phi \in \Upsilon$, if $/\phi/ \neq \mathcal{A}_\Upsilon$ then $f_{/\phi/}(u) = h_\phi(\varsigma(u))$.
   • $f_{\mathcal{A}_\Upsilon}(u) = u$.

5. $V(p) = /p/$.

Before proceeding, we must prove that the f_X's are well defined. That is, in order for for definition 5.2.8 to be a coherent definition, the following proposition must hold:

Proposition 5.2.1:
For any ϕ, $\psi \in \Upsilon$, $/\phi/ = /\psi/ \Rightarrow f_{/\phi/}(u) = f_{/\psi/(u)}$.

Proof. Take then ϕ and $\psi \in \Upsilon$ such that $/\phi/ = /\psi/$. That is (by definition 5.2.2), $\forall \Gamma \in \mathcal{A}_\Upsilon$, $\phi \in \Gamma \Leftrightarrow \psi \in \Gamma$. We must show that for all $u \in \mathcal{U}_\Upsilon$, $f_{/\phi/}(u) = f_{/\psi/}(u)$. We consider the two possible cases.

- Firstly, suppose that $/\phi/ = \mathcal{A}_\Upsilon$. Then by hypothesis $/\psi/ = \mathcal{A}_\Upsilon$. So (by definition 5.2.8), $f_{/\phi/}(u) = f_{/\psi/}(u) = u$.
- Otherwise, neither $/\phi/$ nor $/\psi/$ is $\mathcal{A}_\Upsilon$. Then (by definition 5.2.8), $f_{/\phi/}(u) = h_\phi(\varsigma(u))$ and $f_{/\psi/}(u) = h_\psi(\varsigma(u))$. So it remains to show that $h_\phi(\varsigma(u)) = h_\psi(\varsigma(u))$. By hypothesis, $/\phi/ = /\psi/$, i.e. $\forall \Gamma \in \mathcal{A}_\Upsilon, \phi \in \Gamma \Leftrightarrow \psi \in \Gamma$. That is, $\{\phi, \neg\psi\}$ is inconsistent and $\{\neg\phi, \psi\}$ is inconsistent. So $\vdash \phi \supset \psi$ and $\vdash \psi \supset \phi$, i.e. $\vdash \phi \equiv \psi$. Then by *Sub*, $\vdash [*\phi]\chi \equiv [*\psi]\chi$ for any χ.

  In that case, take any $\delta \in h_\phi(\varsigma(u))$. Then (definition 5.2.6) $[*\phi]\delta \in \varsigma(u)$. So instantiating χ as δ, we have $\varsigma \vdash [*\psi]\delta$, so that as required $\delta \in h_\psi(\varsigma(u))$.

☻

Notice that revision by $[*\top]$ keeps one at the same point, so that it will not necessarily terminate at a point in $\mathcal{U}_\Upsilon$, unlike revision by any other *non-equivalent* formula. (The only formulae revision by which the definition of the orthodox model allows to remain at the same world are those which are equivalent, in DDL_C to $\top$: Suppose that there is some formula $\phi \in \Upsilon$ such that $/\phi/ = \mathcal{A}_\Upsilon$; then there is no $\Gamma \in \mathcal{A}_\Upsilon$ such that $\Gamma \vDash \neg\phi$; so $\vdash_{DDL_C} \phi$.)

Now the orthodox model is almost what we need. In it, as we will show, $j_c \subseteq (j_b)^+$. However, the reverse inclusion does not in general hold. Therefore in the next section we will define a new model which is exactly like the orthodox model except that j_c is made 'regular'.

Remark 5.2.4:
Observe that the way in which we have defined the $j_\Diamond$'s is such that $j_\Diamond(u) \subseteq \mathcal{B}_\Upsilon$.

This is important, as it means that the images of these functions can be defined by a finite set of formulae, and thus by a formula (this is a time when the p_∞ will be used):

Lemma 5.2.2 (Finite definability):
For any $X \subseteq \mathcal{B}_\Upsilon$ there is a formula $\ulcorner X \urcorner$ such that $\forall u \in \mathcal{U}_\Upsilon$, $u \in X \Leftrightarrow u \vdash \ulcorner X \urcorner$. We write $\ulcorner \Gamma \urcorner$ for $\ulcorner \{\Gamma\} \urcorner$.

Proof. For each $\Gamma \in X$, we define $\ulcorner \Gamma \urcorner = \neg p_\infty \wedge \bigwedge \Gamma$. Γ is finite, so this is a formula.

Now suppose $u \vdash \ulcorner \Gamma \urcorner$. Then $u \in \mathcal{B}_\Upsilon$, as $\ulcorner X \urcorner \vdash \neg p_\infty$. And so for any $\Delta \in \mathcal{U}_\Upsilon$, $\Delta \vdash \ulcorner \Gamma \urcorner \Leftrightarrow \Delta = \Gamma$, because these are maximally consistent subsets: If $\Delta \neq \Gamma$, then $\exists \delta \in \Delta, \delta \notin \Gamma$, so that $\Gamma \vdash \neg\delta$, but then $\Delta \vdash \neg\delta$, so $\Delta \vdash \bot$.

Furthermore, since $\mathcal{B}_\Upsilon$ is finite, X will be finite, and so we can define $\ulcorner X \urcorner = \bigvee_{\Gamma \in X} \ulcorner \Gamma \urcorner$.

☻

Lemma 5.2.3 (Inclusion lemma):
In the orthodox model, $j_c \subseteq j_b^+$.

Proof. Take any $u \in \mathcal{U}_\Upsilon$.

Then if we take any $\Delta \in j_c(u)$, we know (remark 5.2.4) that $\Delta \in \mathcal{B}_\Upsilon$. We must show that $\Delta \in j_b^+(u)$.

By remark 5.2.4 again, and by lemma 5.2.2, there is a formula $\ulcorner j_b^+(u) \urcorner$ such that $\forall v \in \mathcal{U}_\Upsilon$, $v \vdash \ulcorner j_b^+(u) \urcorner \Leftrightarrow v \in j_b^+(u)$. If we can show that $u \vdash C \ulcorner j_b^+(u) \urcorner$, we are done.

We can do that if we can show that $\vdash \ulcorner j_b^+(u) \urcorner \supset B \ulcorner j_b^+(u) \urcorner$. For then by C-necessitation, $\vdash C(\ulcorner j_b^+(u) \urcorner \supset B \ulcorner j_b^+(u) \urcorner)$. So by axiom (C2), (*) $\vdash B \ulcorner j_b^+(u) \urcorner \supset C \ulcorner j_b^+(u) \urcorner$. And since $u \vdash B \ulcorner j_b^+(u) \urcorner$, then $u \vdash C \ulcorner j_b^+(u) \urcorner$ as required.. (Suppose $u \nvdash \neg b \neg \ulcorner j_b + (u) \urcorner$; then $\exists \Theta \in j_b(u) - j_b^+(u)$; but by definition 4.2.4, $j_b \subseteq j_b+$, so $j_b - j_b^+ = \varnothing$.)

So it remains to show that $\vdash \ulcorner j_b^+(u) \urcorner \supset B \ulcorner j_b^+(u) \urcorner$. Intuitively this holds, for if you can be reached by a b-chain from u, then any point that you can reach by a single b-step will also be reachable by a b-chain from u. Figure 5.2 illustrates the situation.

Suppose then that $\vdash \ulcorner j_b^+(u) \urcorner \supset B \ulcorner j_b^+(u) \urcorner$ does not hold and derive a (more formal) contradiction: If $\ulcorner j_b^+(u) \urcorner \wedge \neg B \ulcorner j_b^+(u) \urcorner \nvdash \bot$, then there's some $\Gamma \in j_b^+(u)$ such that $\Gamma \nvdash B \ulcorner j_b^+(u) \urcorner$; in which case there is some $\Delta \in \mathcal{B}_\Upsilon - j_b^+(u)$ such that $\Delta \in j_b(\Gamma)$, so $\Delta \in j_b^+(u)$, which is a contradiction. ☻

That done, we move to the final model, in which j_c will be made to conform.

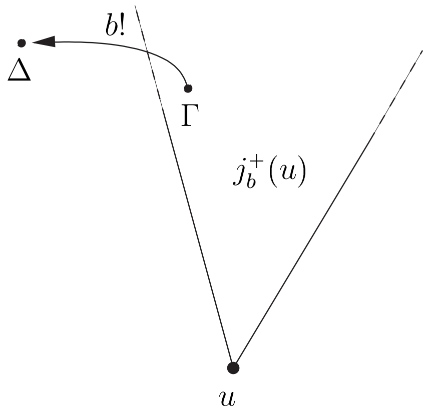


Figure 5.2: Lemma 5.2.3

5.2.3 Sanctioned Model

Definition 5.2.9 (Sanctioned Model over Υ):
The *Sanctioned model* over Υ is the tuple $\mathcal{M}_\Upsilon$:

$$\langle \mathcal{U}_\Upsilon, \tau, \{f_\Diamond\}_{\Diamond \in \{b,c\}}, \{f_X\}_{X \in \tau}, V \rangle,$$

where

- $f_b = j_b$;
- $f_c = f_b^+$;

(And $\mathcal{U}_\Upsilon$, τ, the f_X's, and V's are as above, in definition 5.2.8.)

Now we will prove Υ-restricted versions the standard existence and truth lemmata for the Sanctioned models. The, in §5.4, we will prove that the Sanctioned model is a DDL_C model.

5.3 Completeness

In the following proof, on a number of occasions we will want to appeal to the existence of a certain maximally consistent subset of Υ; thus we provide the following lemma:

Lemma 5.3.1 (Construction lemma):
For $\Diamond \in \{b, c, [*\psi]\}$: Given any $u \in \mathcal{U}_\Upsilon$ such that $\Diamond\phi \in u$, there exists a $\Delta \in \mathcal{A}_\Upsilon$ such that $\Delta \vdash \phi$ and $u \nvdash \neg\Diamond\ulcorner\Delta\urcorner$.

Proof. Consider the following set:

$$\Lambda = \{\Gamma \subset \Upsilon | u \nvdash \neg\Diamond\ulcorner\Gamma\urcorner\}.$$

Enumerate all of the formulae $\nu_0 \dots \nu_k \in \Upsilon$ such that for each ν_i, there is no formula ψ such that $\nu_i = \neg\psi$. Then we inductively define a $\Delta \in \mathcal{A}_\Upsilon$ that will be an element of Λ :

(0) $\Delta_0 = \{\nu_0\}$

(I) $\Delta_{n+1} = \begin{cases} \Delta_n \cup \{\nu_{n+1}\} & \text{if } u \nvdash \neg\Diamond(\nu_{n+1} \wedge \ulcorner\Delta_n\urcorner) \\ \Delta_n \cup \{\neg\nu_{n+1}\} & \text{otherwise} \end{cases}$

Clearly Δ_0 is in Λ; furthermore $\Delta_k \in \mathcal{A}_\Upsilon$: the process of construction creates a *maximally* consistent subset of Υ. Thus it only remains to show that the step (I) preserves membership of Λ. To do this, suppose that $\Delta_n \in \Lambda$. Then $\forall\chi, (u \nvdash \neg\Diamond(\chi \wedge \ulcorner\Delta_n\urcorner)$ *or* $u \nvdash \neg\Diamond(\neg\chi\ulcorner\Delta_n\urcorner))$: For suppose this were not the case; then

$u \vdash \neg\Diamond(\chi \wedge \ulcorner\Delta_n\urcorner) \wedge \neg\Diamond(\neg\chi \wedge \ulcorner\Delta_n\urcorner)$, i.e.
$u \vdash \Box\neg(\chi \wedge \ulcorner\Delta_n\urcorner) \vee \Box\neg(\neg\chi \wedge \ulcorner\Delta_n\urcorner)$, so
$u \vdash \Box(\neg(\chi \wedge \ulcorner\Delta_n\urcorner) \vee \neg(\neg\chi \wedge \ulcorner\Delta_n\urcorner))$, i.e.
$u \vdash \Box((\neg\chi \vee \neg\ulcorner\Delta_n\urcorner) \vee (\chi \vee \neg\ulcorner\Delta_n\urcorner))$, i.e.
$u \vdash \Box\neg\ulcorner\Delta_n\urcorner$.

But this last line contradicts the fact that $\Delta_n \in \Lambda$.
This means that either $\Delta_n \cup \{\nu_{n+1}\} \in \Lambda$ xor $\Delta_n \cup \{\neg\}\nu_{n+1} \in \Lambda$, and the construction (I) choose the one that is in Λ, so that we get $\Delta_{n+1} \in \Lambda$.

Then Δ_k is the construction we were looking for. ☻

It is in fact the following immediate corollary of the construction lemma that we will use in the two existence lemmata which proceed it.

Corollary 5.3.1 (Belief-State construction):
For $\Diamond \in \{b, c\}$, for $\Diamond\phi \in \Upsilon$: Given any $u \in \mathcal{U}_\Upsilon$ such that $\Diamond\phi \in u$, there exists a $\Delta' \in j_\Diamond(u)$ such that $\Delta' \vdash \phi$.

Proof. Let Δ be that constructed by the construction lemma, i.e. $\Delta \in \mathcal{A}_\Upsilon$ and $\Delta \vdash \phi$ and $u \nvdash \neg\Diamond\ulcorner\Delta\urcorner$. Then let $\Delta' = \Delta \cup \{\neg p_\infty\}$. ☻

Lemma 5.3.2 (*B*-existence):
$\forall b\phi \in \Upsilon, \forall u \in U_\Upsilon, b\phi \in u \Leftrightarrow f_b(u) \cap /\phi/ \neq \emptyset$.

Proof. $\Rightarrow$: Hypothesis: $b\phi \in u$. By applying the construction corollary there is a $\Delta \in j_b(u)$ such that $\Delta \vdash \phi$. But $f_b = j_b$.

$\Leftarrow$: Hypotheses: (1) $b\phi \in \Upsilon$; (2) $\phi \in \Delta$; and (3) $\Delta \in f_b(u)$. By (3), $u \nvdash \neg b\ulcorner\Delta\urcorner$, so by (2), $u \nvdash \neg b\phi$; so by (1), $u \vdash b\phi$.

☻

Lemma 5.3.3 (*C*-existence):
$\forall c\phi \in \Upsilon, \forall u \in U_\Upsilon, c\phi \in u \Leftrightarrow f_c(u) \cap /\phi/ \neq \varnothing$.

Proof. $\Rightarrow$: Hypothesis: $c\phi \in u$. We must demonstrate the existence of $\Delta' \in f_c(u)$ with $\phi \in \Delta'$. Here we use the Inclusion Lemma (lemma 5.2.3), which tells us that if $\Delta \in j_c(u)$ then $\Delta \in f_b^+(u)$, and so since $f_c = (f_b)^+$, it will suffice to construct a Δ such that $\Delta \in j_c(u)$ and with $\phi \in \Delta$; this is given to us again by the construction corollary.

$\Leftarrow$: Hypotheses: (1) $c\phi \in \Upsilon$; and for some for some Δ, (2) $\phi \in \Delta$ and (3) $\Delta \in f_c(u)$. By (3), there is some finite sequence $\Theta_0 \ldots \Theta_l$ with $\Theta_i \in f_b(\Theta_{i+1})$ for $i < l$ such that $u = \Theta_l$ and $\Delta \in f_b(\Theta_0)$. (See figure 5.3.) By induction on l we show that $c\phi \in \Theta_l = u$. We will take as our inductive hypothesis that for all such chains (of the type depicted in figure 5.3) of length k, $c\phi$ is an element of the last (right-most in the diagram) member of the chain.

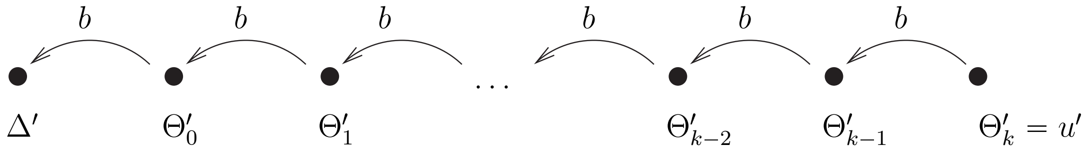


Figure 5.3: A chain of the type the induction is over in the *C*-existence lemma.

(0) Where $k = 0$, $\Delta \in f_b(u)$, so that $u \nvdash \neg b \bigwedge \Delta$. So by (2), $u \nvdash \neg b\phi$. By axiom (C1), $\vdash \neg c\phi \supset \neg b\phi$, so $u \nvdash \neg c\phi$. Therefore by (1), $u \vdash c\phi$.

(I) $k = (n+1)$. Inductive hypothesis: $c\phi \in \Theta_i$ for every $i < k$. Then in particular $c\phi \in \Theta_n$, and so, because $\Theta_n \in f_b(\Theta_k)$, $\Theta_k \nvdash \neg bc\phi$. But by by axiom (C1), $\vdash \neg c\phi \supset \neg bc\phi$; so $\Theta_k \nvdash \neg c\phi$. So by (1), $c\phi \in \Theta_k$.
[4]

☻

Lemma 5.3.4 ($*\phi$-existence):
$\forall [*\phi]\psi \in \Upsilon, \forall u \in U_\Upsilon, [*\phi]\psi \in u \Leftrightarrow \psi \in f_{/\phi/}(u)$.

[4]The Fischer-Ladner condition ($c\phi \in \Upsilon \Rightarrow bc\phi \in \Upsilon$) is used in [19] for the completeness of *PDL*, and it is suggested in the proof in [9] (pp 240–246) that it is needed in a *PDL* completeness proof which uses similar techniques to ours. Can we do without it altogether on our proof?

Proof. We distinguish two cases. One in which $\vdash \phi$, and a second where $\neg\phi$ is consistent.

1. Suppose that $\vdash \phi$. I.e. $\vdash \phi \equiv \top$. Then by *Sub*, $\vdash [*\phi]\psi \equiv [*\top]\psi$. And by (R7) $\vdash [*\top]\psi \equiv \psi$, so $\vdash [*\phi]\psi \equiv \psi$. Then in particular we have $[*\psi]\psi \in u \Leftrightarrow \psi \in u$.

   The supposition that $\vdash \phi$ also entails that $f_{/\phi/}(u) = f_{/\top/}(u) = u$ (definition 5.2.8). Then we indeed have $[*\phi]\psi \in u \Leftrightarrow \psi \in f_{/\phi/}(u)$.

2. Suppose on the contrary, as a hypothesis, that $\neg\phi$ is consistent. Then $/\phi/ \neq /\top/$, meaning that we use the other part of the definition of $f_{/\phi/}(u)$ from definition 5.2.8, namely that

   (1) $f_{/\phi/}(u) = h_\phi(\varsigma(u))$.

   Now, definition 5.2.6 gives us

   (2) $[*\phi]\psi \in \varsigma(u) \Leftrightarrow \psi \in h_\phi(\varsigma(u))$.

   And definition 5.2.7 gives us

   (3) $[*\phi]\psi \in u \Leftrightarrow [*\phi]\psi \in \varsigma(u)$.

   The result comes immediately from putting (1)–(3) together.

☻

Lemma 5.3.5 (Truth Lemma):
$\forall\phi \in \Upsilon, \forall u \in U_\Upsilon, \phi \in u \Leftrightarrow u \vDash_{\mathfrak{M}_\Upsilon} \phi$.

Proof. By induction on ϕ.

The base case is immediate: $p \in u \Leftrightarrow u \in V(p) \Leftrightarrow u \vDash p$.

Assume that the lemma holds for formulae less complex that χ, and consider:

$\neg$: Suppose $\chi = \neg\phi$

$$\begin{array}{rcll} \neg\phi \in u & \Leftrightarrow & \phi \notin u & (\because \chi \in \Upsilon) \\ & \Leftrightarrow & u \nvDash \phi & \text{(inductive hypothesis)} \\ & \Leftrightarrow & u \vDash \neg\phi & \end{array}$$

$\wedge$: Suppose $\chi = \phi \wedge \psi$.

$$\begin{array}{rcl} \phi \wedge \psi \in u & \Leftrightarrow & \phi \in u \;\&\; \psi \in u \\ & \Leftrightarrow & u \vDash \phi \;\&\; u \vDash \psi \\ & \Leftrightarrow & u \vDash \phi \wedge \psi \end{array}$$

$\Diamond$: Suppose $\chi = \Diamond\phi$, where $\Diamond \in \{b, c\}$.

$$\begin{array}{lll} \Diamond\phi \in u & \Leftrightarrow & f_c(u) \cap /\phi/ \neq \varnothing & (\Box\text{-existence lemma}) \\ & \Leftrightarrow & \exists \Delta \in f_c(u) : \Delta \vDash \phi & (\text{inductive hypothesis}) \\ & \Leftrightarrow & u \vDash \Diamond\phi \end{array}$$

$\langle *\phi \rangle$: Suppose $\chi = [*\phi]\psi$.

We know that $\phi \in \Upsilon$ by the closure condition Π. Therefore $/\phi/ \in \tau$. So the satisfaction definition for $[*\phi]$ (see definition 4.2.3) gives us:

$$\begin{array}{lll} [*\phi]\psi \in u & \Leftrightarrow & f_{/\phi/}(u) \vdash \psi & ([*\phi]\text{-existence lemma}) \\ & \Leftrightarrow & u \vDash [*\phi]\psi \end{array}$$

☻

Given the standard existence and truth lemmata that we have just proved, it is now a standard step to prove weak completeness of the logic with respect to Sanctioned models. However, we want to prove weak completeness with respect to DDL_C models, so there is one big step left to make, which amounts to showing that we do indeed have selection functions (cf. definition 4.2.2):

5.4 Sanctioned models are DDL_C models

Proposition 5.4.1:
The Sanctioned model over Υ is a DDL_C model.

Proof. (a) Immediate from definition 5.2.8.

(b) $\Upsilon^{\curlyvee} = \{/\phi/ | \phi \in \Upsilon^{\curlyvee}\}$ is a Boolean algebra over $\mathcal{U}_\Upsilon$: By construction.

(c) Immediate from definition 5.2.8.

(d) $f_b(f_{/\phi/}(u))$ is a selection function over $\mathcal{U}_\Upsilon$:

(1) $f_b(f_{/\phi/}(u)) \subseteq /\phi/$: Take any $\Delta \in f_b(f_{/\phi/}(u))$. Suppose towards contradiction that $\Delta \nvDash \phi$. Then $\Delta \vdash \neg\phi$ (lemma 5.2.1). So $f_{/\phi/}(u) \vdash b\neg\phi$. Then $\varsigma(u) \vdash [*\phi]b\neg\phi$, giving us a contradiction with the combination of axiom (R1) and the consistency of $\varsigma(u)$.

(2) $/\phi/ \subseteq /\psi/ \Rightarrow (f_b(f_{/\phi/}(u)) \neq \varnothing \Rightarrow f_b(f_{/\psi/}(u)) \neq \varnothing))$. Assume that $/\phi/ \subseteq /\psi/$. Then $\vdash \phi \supset \psi$; otherwise $\nvdash \phi \wedge \neg\psi$, so there would be some consistent set containing ϕ and $\neg\psi$; but then there would be a member of $\mathcal{B}_\Upsilon$ containing ϕ and not containing ψ, contracting the assumption.

Then by rule (KN), $\vdash [*\phi \wedge \neg\psi]\bot$; so by axiom (R2), (*) $\vdash [*\psi]B\bot \supset [*\phi]B\bot$.

Now assume that $f_b(f_{/\phi/}(u)) \neq \varnothing$. Then there is some Δ such that $b \bigwedge(\Delta - \{\neg p_\infty\} \in f_{/\phi/}(u)$. So $\varsigma(u) \vdash [*\phi]b\top$. So by (*), $\varsigma(u) \vdash [*\psi]b\top$. So $f_{/\psi/}(u) \vdash b\top$, so there is some point in $f_b(f_{/\phi/}(u))$ as required.

(3) $/\phi/ \subseteq /\psi/ \Rightarrow (/\phi/ \cap f_b(f_{/\psi/}(u)) \neq \varnothing) \Rightarrow f_b(f_{/\phi/}(u)) = /\phi/ \cap f_b(f_{/\psi/}(u))$.
Assume as a hypothesis that $/\phi/ \subseteq /\psi/$. Then, as we have already seen, $\vdash \phi \supset \phi$.

So we can conclude that $\vdash (\psi \wedge \phi) \equiv \phi$.

Then assume further than $/\phi/ \cap f_b(f_{/\psi/}(u)) \neq \varnothing$. Now we will show both halves of the inclusion required to show that

$$f_b(f_{/\phi/}(u)) = /\phi/ \cap f_b(f_{/\psi/}(u)) :$$

$\subseteq$: We have already seen that $f_b(f_{/\phi/}(u)) \subseteq /\phi/$, so it remains to show that $f_b(f_{/\phi/}(u)) \subseteq f_b(f_{/\psi/}(u))$. Well, we know that, since $\varsigma(u)$ is a maximally consistent extension of w[5], $\varsigma(u) \vdash [*\psi]b\phi$. So, using the fact that $\vdash \phi \supset \psi$ along with axiom (R5) and rule (RE), we get $\varsigma(u) \vdash [*\psi]B(\phi \supset \gamma) \supset [*\phi]B\gamma$ for any γ.

Now for γ we are going to take $\ulcorner f_b(f_{/\psi/}(u))\urcorner$.

Clearly $\varsigma(u) \vdash [*\psi]B\ulcorner f_b(f_{/\psi/}(u))\urcorner$.

(Otherwise $f_{/\psi/}(u) \nvdash \neg b\neg\ulcorner f_b(f_{/\psi/}(u))\urcorner$, i.e. there's some point Δ such that $\Delta \notin f_b(f_{/\psi/}(u))$ and $f_{/\psi/}(u) \vdash b\ulcorner\Delta\urcorner$, so $\Delta \in f_b(f_{/\psi/}(u))$!) Then $\varsigma(u) \vdash [*\psi]B(\phi \supset \ulcorner f_b(f_{/\psi/}(u))\urcorner)$. So $\varsigma(u) \vdash [*\phi]B\ulcorner f_b(f_{/\phi/}(u))\urcorner$. Then any $\Gamma \in f_b(f_{/\phi}(u))$ will be such that $\Gamma \in f_b(f_{/\psi}(u))$.

$\supseteq$: Consider the following instantiation of (R4): $\vdash [*\phi]B\ulcorner f_b(f_{/\phi/}(u))\urcorner \supset [*\psi]B(\phi \supset \ulcorner f_b(f_{/\phi/}(u))\urcorner)$. And since $\varsigma(u) \vdash [*\phi]B\ulcorner f_b(f_{/\phi/}(u))\urcorner$, $\varsigma(u) \vdash [*\psi]B(\phi \supset \ulcorner f_b(f_{/\phi/}(u))\urcorner)$. So $f_{/\phi/}(u) \vdash B(\neg\phi \vee \ulcorner f_b(f_{/\phi/}(u))\urcorner)$. Then consider some $\Gamma \in /\phi/$, i.e such that $\Gamma \vdash \phi$. In such a case, $f_{/\psi/}(u) \vdash b(\bigwedge\Gamma \wedge \phi)$. But then $f_{/\psi/} \vdash b(\bigwedge\Gamma \wedge \ulcorner f_b(f_{/\phi/}(u))\urcorner)$, so $\Gamma \vdash \ulcorner f_b(f_{/\phi/}(u))\urcorner$, i.e. $\Gamma \in f_b(f_{/\phi/}(u))$.

(e–f) Immediate from the fact (see definition 5.2.8) that $f_{/\top/}(u) = u$.

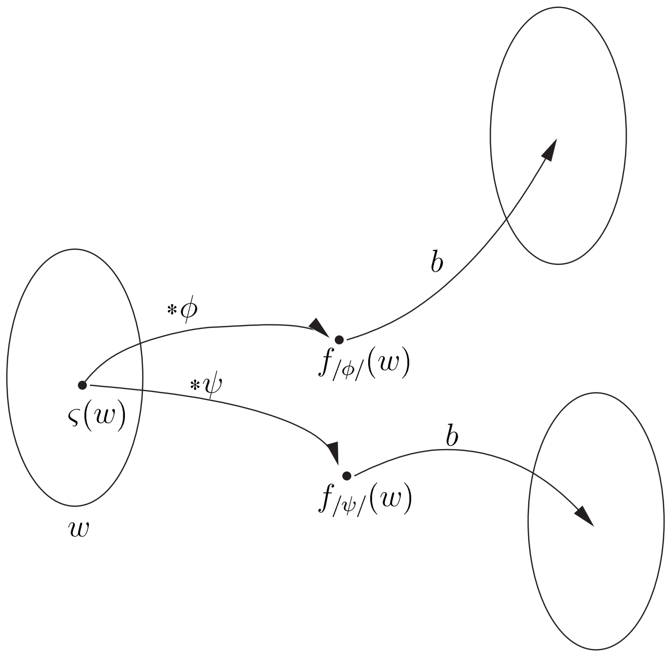
(g) $f_c = f_b^+$: Immediate from definition 5.2.9.

☻

Now that we have shown that, the remainder is plain sailing:

Theorem 5.4.1 (Weak Completeness of DDL_C – Proposition 4.2.1):
Given any consistent formula ξ, there is a DDL_C-model $\mathfrak{M}$ containing a point u such that $u \vDash_{\mathfrak{M}} \xi$.

[5]When we talk of maximality we refer to maximality in the language *not* including p_∞.



Proof. Take $\Upsilon = \Pi(\{\xi\})$, then the Sanctioned model over Υ will contain a point u such that $\xi \in u$, which by the truth lemma means that $u \vDash \xi$. ☻

6 Algebraic Approach

6.1 Introduction

Baltag and Sadrzadeh [5] present a different way of reasoning about the dynamics of information in a multi-agent context. The semantics of their system is algebraic, and they show it to be more general than dynamic epistemic logic. The proof theory for the system is not a Hilbert-style axiom system, as with DEL, but is a Gentzen-style sequent calculus, containing features from linear logic. However, that proof system is not complete for the fragment that includes the revision operator[1], and there are good reasons, due to the algebraic approach over a more standard modal one, why for example techniques from the last chapter cannot be used to provide a syntactic characterisation of the operators.

We summarise the algebraic semantics they present, and show how it connects to dynamic epistemic logic. We investigate the connections with AGM theory, and this involves considering what propositional meaning we might attribute to a given action. We suggest that the last two AGM postulates are not respected by the axioms proposed in [5], and finally look at why providing a complete proof system for the revision operator, even with only these limited axioms, is not a trivial problem.

6.2 Expansion

6.2.1 Algebraic Semantics

The algebraic structure presented in [26] for reasoning about information change is general in the sense that the lattices used to represent propositions and actions need not be Boolean; reasons for adopting such an approach are tentatively argued for in [5]: "there is no meaningful notion of 'negation of a theory'; similarly, there is no action that can be called 'the negation' of action q." Whether or not we endorse this view[2], it is interesting

[1] [26] contains a completeness proof for the non-revision fragment.

[2] The first half of their remark is certainly prima facie appealing. We could remark against the second that there is no reason to take an action q's sitting opposite $\neg q$ in a lattice of actions to mean that $\neg q$ is the negation of the action q; rather, for example in the case of the porting of DEL to the algebraic setting, the $\neg q$, the negation of an action q (which is really a set of atomic actions) is the action which is everything that q is not; i.e. the set of actions which are not in q.

to observe that so much can be done in a less restrictive setting, and we can (and will) consider only Boolean lattices when the need arises.

Definition 6.2.1 (Lattice of Propositions):
A *lattice of propositions* is just a complete lattice $\mathcal{M} = \langle M, \vee, \wedge \rangle$, where M is a set.

The order[3] in the lattice is supposed to represent strength of information: the smaller the element, the stronger its informational content. If $m \leqslant m'$ then m entails m'. A descending path from $\top$ towards $\perp$ represents an increase in information. If we take theories in the *AGM* sense to be logically closed sets of formulae, then the lattice of theories over the language $\mathcal{L}$ would be $(2^{\mathcal{L}})^{\vdash}$, the logically closed sets of $\mathcal{L}$-formulae, and set intersection ($\cap$) as meet ($\vee$), and $\cup^{\vdash}$ as join ($\wedge$), where $A \cup^{\vdash} B = (A \cup B)^{\vdash}$. The ordering by information strength is such that we will read $m \leqslant m'$ as m entails m'. From the *AGM* perspective, a move from m' to m, where $m \leqslant m'$, is a belief expansion.

Definition 6.2.2 (Lattice of Actions):
A *lattice of actions* is a complete lattice with an additional (order-preserving) monoidal structure $\mathcal{Q} = \langle Q, \triangledown, \vartriangle, \bullet, 1 \rangle$.

Again, a member of the lattice is more determined the lower it is. If an action q was an announcement that ϕ is true, and an action q' an announcement that ψ is true, then the action $q \vee q'$ would be an action that was either an announcement of ϕ or an announcement of ψ. (Which is not in general the same as an announcement that $\phi \vee \psi$ is true.) The unit 1 in the lattice is to be thought of as the 'empty' action, 'skip' in computer science jargon. The monoid operation is concatenation, so that $q \bullet q'$ in the example above would be the action that is an announcement of ϕ followed by an announcement of ψ. The next definition connects these two structures, so that we are able to talk about the effect of an action.

Definition 6.2.3 (Proposition-Action System):
A *proposition-action system* is a tuple $\langle \mathcal{M}, \mathcal{Q}, \otimes \rangle$ consisting of a lattice of propositions and one of actions, and with $\otimes : M \times Q \to M$ satisfying the following conditions:

1. $(\bigvee_{i \in I} m_i) \otimes (\bigvee_{j \in J} q_j) = \bigvee_{i \in I} \bigvee_{j \in J} (m_i \otimes q_j)$
2. $m \otimes 1 = m$
3. $m \otimes (q \bullet q') = (m \otimes q) \otimes q'$

$m \otimes q$ is to be understood as the application of the action q to the proposition m: If at some moment m holds, then after q happens, $m \otimes q$ will hold. In this context, the conditions should make good sense. For example, the last one says that the effect of

[3] recall that $m \leqslant m'$ iff $m' = m \vee m'$

doing (q then q') is doing q, then doing q'. We will call such operators '(algebraic) expansion functions' – $m \otimes q$ is the expansion of the current situation m by adding to it the occurrence of q.

Definition 6.2.4 (Local View Expansion Function):
Agent i's *local view expansion function* f_i is a pair $\langle f_i^M : M \to M, f_i^Q : Q \to Q \rangle$, preserving arbitrary joins, and satisfying the following conditions:

1. $1 \trianglelefteq f_i^Q(q)$
2. $f_i^Q(a \bullet b) \trianglelefteq f_i^Q(a) \bullet f_i^Q(b)$
3. $f_i^M(m \otimes a) \leqslant f_i^M(m) \otimes f_i^Q(a)$

A local view function is much like those from relational semantics: The idea will be that if the situation is m, then agent i will think that it is $f_i^M(m)$. We should really be wary of saying that a situation ever *is* indeterminate in the way in which some of the members of the lattice of propositions are. Furthermore, instead of saying that the agent thinks that the situation is a proposition $f_i^M(m)$, we should say that the agent thinks that the actual (i.e. fully determined) situation is a member of $f_i^M(m)$. To describe 'real-world', fully determined states we will use 'atoms':

Definition 6.2.5 (Atom):
$x \in \langle X, \leqslant, \perp \rangle$ is *atomic* iff$_{df}$ $\forall y \neq \perp (y \leqslant x \Rightarrow x \leqslant y)$. We denote the set of atoms of X as $Atm(X)$.

The non-atoms represent indeterminacy, which, at least in the modeling of social situations, we might wish to avoid ascribing to the real world.

Definition 6.2.6 (n-agent Algebraic Expansion Model):
A proposition-action system supplemented with a set $\{f_i\}_{i \in n}$ of local view expansion functions is an *n-agent algebraic model*.

In order to reason about beliefs in this system, we need a belief modality. Where we had a B_i operator in our language corresponding to the local view function in the relational setting, here we will introduce another such operator into the algebra:

$$B_i(m) = \bigvee \{m' | f_i(m') \leqslant m\}$$

The connection between f_i and B_i is such that we have the adjunction

$$\frac{f_i(m) \leqslant m'}{m \leqslant B_i(m')}$$

Thus where f_i preserved arbitrary joins, B_i preserves arbitrary meets.

So far there is a fairly clear parallel with DEL. In section 6.3.1 we will see one simple way of adding belief revision to the algebraic setting, and we will go on in the following sections to compare this to AGM. Before that, we will first sketch how the parallel with DEL can be made more precise.

6.2.2 DEL algebraically

[26] provides a representation theorem for dynamic epistemic models. By closing the action model under composition, and the state model under update (forming a union of disjoint models), it is possible to generate an n-agent algebraic model.

Definition 6.2.7 (Full DEL model):
A DEL model $\langle \mathcal{S}, \mathcal{A} \rangle$ is *full* iff$_{df}$

1. it is closed under update: $\mathcal{S} \boxtimes \mathcal{A} \subseteq \mathcal{S}$;
2. it is closed under concatenation of actions; i.e. $\forall q, q' \subseteq \mathcal{A}$, $\exists q'' \subseteq \mathcal{A} : \forall m \subseteq \mathcal{S}, m \boxtimes q'' = (m \boxtimes q) \boxtimes q'$;
3. and it contains a 'skip' action: $\exists q \subseteq \mathcal{A} : \forall m \subseteq \mathcal{S}, m \boxtimes q = \mathcal{S}$.

Any DEL model will generate a full DEL model containing it, by closure under $\boxtimes$ and composition. The full model will in general be larger than the DEL model: it will contain states that represent the 'effect' of each action on each state. We will look in section 7.2 at the connection between these larger models and the history-based ('epistemic temporal') models of [24]. First though, we finish our summary of the representation theorem:

Definition 6.2.8 (Concrete Epistemic System):
The *concrete epistemic system* induced by the full DEL model $\langle \mathcal{S}, \mathcal{A} \rangle$ is$_{df}$ $\langle 2^{\mathcal{S}}, 2^{\mathcal{A}} \rangle$.

A concrete epistemic system, as a pair of power set lattices, and when supplemented with the appropriate functions ($\otimes$ is $\boxtimes$; $\bullet$ is the concatenation of actions; 1 is the 'skip' action; and the local view functions are those of the full DEL model lifted to its power set: $f_a(\bigcup_i x_i) = \bigcup_i f_a(x_i)$) forms an n-agent algebraic expansion model.[4]

We will discuss some aspects of this representation theorem a little further in §7.2. For now, revision:

[4] Details are more extensive in [26].

6.3 Revision

The $\otimes$ operator expressed expansion, in effect coinciding with the same operator DEL. How can we introduce revision?

6.3.1 Algebraic Axiomatic Definition

In [5] another operator $\circledast$ is introduced to the algebra which is meant to represent *revision*, where $\otimes$ represented expansion.

Definition 6.3.1 (Algebraic Propositional Revision Function):
Given a proposition-action system $\langle \mathcal{M}, \mathcal{Q}, \otimes \rangle$, $\circledast : M \times Q \to M$ is an *algebraic propositional revision function* precisely when it satisfies the following conditions:

($\circledast$1) $\exists m' m \circledast a = m' \otimes q$[5]

($\circledast$2) $m \otimes a \neq \bot \Rightarrow m \circledast a = m \otimes a$

($\circledast$3) $a \neq \bot \Rightarrow m \circledast a \neq \bot$

($\circledast$4) $m \otimes a \leqslant m \circledast a$

The motivation for introducing such a function is to generalise the AGM notion of revision. The connection is like this: If q is an observation, or announcement, of ϕ, then $m \otimes q$ is to $\Gamma_m \cup \{\phi\}$ what $m \circledast q$ is to $\Gamma_m \dotplus \phi$. Notice for example that the first three conditions appear to be pairwise-equivalent to, or at least to have a very similar flavour as, the first three AGM postulates. These last two statements are vague, and while we make the intuition behind them more explicit below, for example with proposition 6.3.1, we shall see in section 6.3.2 the difficulties involved in making them precise. These difficulties are due to the objections we raised against using the AGM postulates. Firstly, the 'thinness of actions': actions are considered in a more general sense than just as the learning by a single agent of a ground fact, as in AGM. Then 'retention of disbelief': if ψ is consistent with ϕ it will not necessarily hold that if ψ was true, it will be after learning ϕ. And then 'Moore-ish objection': if an agent learns that ϕ, ϕ will not necessarily hold after that action.

These objections, we suggested, have their basis in a view of the process of revision as being an action, or at least being brought about by an action. The elements of the lattice of actions are supposed to represent those actions. In this case, agents can revise their views about what actions have taken place, in addition to just about in what circumstances things happened.

[5]Recall that these formulae are implicitly universally quantified.

Therefore in [5] one finds also analogous conditions for the action revision function, $\circledast^Q : \mathcal{Q} \times \mathcal{Q} \to \mathcal{Q}$:

($\circledast^Q$1) $\exists q'' q \circledast^Q q' = q'' \bullet q'$

($\circledast^Q$2) $q \bullet q' \neq \bot \Rightarrow q \circledast^Q q' = q \bullet q'$

($\circledast^Q$3) $q \neq \bot \Rightarrow q \circledast^Q q' \neq \bot$

($\circledast^Q$4) $q \bullet q' \leqslant q \circledast^Q q'$

The revision function does its work with respect to the agents' local views of events. The agents are able to revise their theories about the world, and also about what actions are taking place (and, given sufficient structure on the model, as we shall see in §7.2 is the case with the algebraic models generated by full DEL models, about what actions have taken place).

Definition 6.3.2 (Local View Revision Function):
Agent i's *local view revision function* f_i is a pair $\langle f_i^M : M \to M, f_i^Q : Q \to Q \rangle$, preserving arbitrary joins, and satisfying the following conditions:

1. $1 \trianglelefteq f_i^Q(q)$
2. $f_i^Q(a \bullet b) \trianglelefteq f_i^Q(a) \circledast^Q f_i^Q(b)$
3. $f_i^M(m \otimes a) \leqslant f_i^M(m) \circledast f_i^Q(a)$

One consequence of this is that each agent will always have a consistent theory, as long as she starts consistent and she does not consider $\bot$ to occur.

Definition 6.3.3 (n-agent Algebraic Revision Model):
A proposition-action system supplemented with an algebraic revision function and a set $\{f_i\}_{i \in n}$ of local view expansion functions is an *n-agent algebraic revision model.*

In order to make precise the connections between the axioms of definition 6.3.1 and the AGM postulates, we must first of all specify when a given action is to qualify as a true statement that ϕ, and thus when $m \circledast q$ is analogous to some $\Gamma \dotplus \phi$. We discuss this problem in the following section. After that, in §6.3.3, we consider a restricted class of algebraic revision models, which are closer to respecting the AGM postulates (in a sense that we make precise).

6.3.2 Actions: Semantics, Pragmatics

In order to make a precise comparison between the AGM postulates and the axiomatic definitions of algebraic revision models (i.e. definition 6.3.1), it will be useful to have some kind of formal 'semantics' for the members of the lattices of actions of those algebraic models. More specifically, we want to give necessary and sufficient conditions C such that a given action $q \in Q$ is a truthful statement that $m \in M$ holds if and only if it satisfies the conditions C.

As we stated in section 2.2.3, the meaning of an action is often taken to be its truth-conditions. That is, the conditions in which an utterance of a sentence is true give the meaning for that sentence. In this setting then, we can talk sensibly about the meaning of an action, and a definition given in [2] is in effect an attempt to do just that. We will see later in this section though, that the other factors – the pragmatic parts of the meaning of an action – are (unsurprisingly) not fully captured by what is effectively a definition of the semantics of an action given in [2]. We will now discuss that definition, just after giving two simple definitions. Firstly, the *kernel* of an action $ker(q)$, which is the set of propositions which rule out the possibility of q taking place.

Definition 6.3.4 (Kernel):

$$ker(q) = \bigvee\{m | m \otimes q = \bot\}$$

The kernel of an action then gives its 'falsity-conditions'. So when an action q has occurred, we know that $ker(q)$ did not hold, because otherwise q could not have occurred.

Remark 6.3.1:

$$ker(q) = [q]\bot$$

The second definition relates the behaviour of an action to its identity:

Definition 6.3.5 (Extensionality):
A proposition-action system $\langle \mathcal{M}, \mathcal{Q}, \otimes \rangle$ is called *extensional* iff$_{df}$ $q \neq q' \Rightarrow \exists m : m \otimes q \neq m \otimes q'$.

Given that we have falsity-conditions, truth-conditions should not be far away. Indeed, [2] provides the following: "[P]ublic announcement of the proposition $m \in M$ is an epistemic action $q \in Q$ for which $f_a(q) = q$ [for each agent a in the group to which the announcement is made/visible], and for which $ker(q)$ has a *Boolean complement* $\neg ker(q)$, satisfying $\neg ker(q) = m$." (p. 6) This is apparently enough for the purposes of that paper, where the muddy children puzzle is worked out in using an algebraic expansion model. We can note, however, that the "is" is not one of definition, for there are several such epistemic actions, even in an extensional system. Furthermore, the condition provided, that an action q must meet in order to be considered as a candidate for being the public announcement of a member of the propositional lattice, is certainly necessary, but is not sufficient.

Let us take the essential part of what we need from the statement cited above from [2]. We will leave out the reference to the local view function f_a, because we are concerned only with an action *being* a truthful statement that a given proposition holds, and not with the way it is perceived by the agents; abstracting away from an action representing an announcement having any audience may seem peculiar, but for the time being we want to consider the action in isolation from any agents who might perceive it. Then let us paraphrase the above quotation from [2]:

($\ast$) $q \in Q$ is a truthful statement that $\neg[q]\bot$.

There can clearly, even in an extensional system, be $q, q' \in Q$ such that $q \neq q'$ and $ker(q) = ker(q')$, and hence $\neg ker(q) = \neg ker(q')$. An example of a simple finite case is given in figure 6.1.

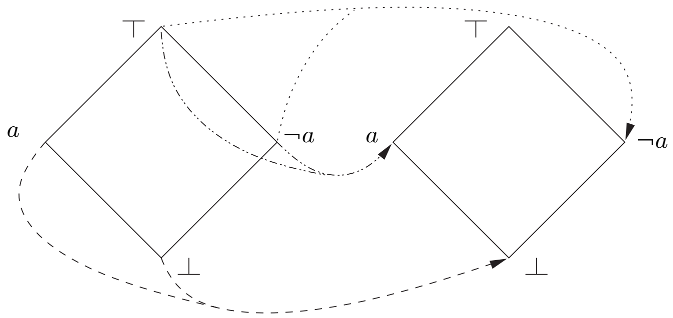


Figure 6.1: Two extensionally different (join-preserving) endomorphisms, on a Boolean lattice, with the same kernel

This reminds us that specifying the kernel of a function does not uniquely determine that function, so that ($\ast$) does not uniquely determine (the extension of) an action which is a truthful statement that m holds. Furthermore, this is actually a problem, leading us to note the insufficiency of ($\ast$): We can see here that in the diagram only one of these morphisms can be considered a truthful statement that m holds. Notice that *we do not know which one*, because to do that we need to know more about the meaning of m.[6]

[6]This does appear to conflict with this loose way of speaking: 'father's announcement q_0 that at least one child is dirty is *given* by setting ($f_A(q_0) = q_0$) &) $ker(q_0) =\downarrow D_{\varnothing}$, where D_C is the proposition that the set of children C are dirty. And the children's simultaneous answer (a *refutation*), q, that they do not know if they are muddy, is similarly given by ($f_A(q) = q$) &) $ker(q) =\downarrow \bigvee_{i \in N} B_i D_i$, where D_i is the proposition that child i is dirty.' ([2], p. 8

What exactly *does* specifying the kernel, the 'co-precondition', fix? We will use the following notation:

Definition 6.3.6 (Co-precondition Semantics of Actions):
An action $q \in Q$ has the *co-precondition semantics* $m \in M$ iff $ker(q) = \neg m$. We write q_m for q in this case.

We easily get the following:

Remark 6.3.2:
$m' \otimes q_m = (m \wedge m') \otimes q_m$

Proof. TST: $m' \otimes q_m = (m' \wedge m) \otimes q_m$
$\neg m \otimes q_m \leqslant \perp$ [Definition of q_m]
$\neg m \otimes q_m \leqslant m''$
$\neg m \leqslant [q_m]m''$
$[q_m]m'' = \neg m \vee [q_m]m''$
$m' \leqslant [q_m]m'' \Leftrightarrow m' \leqslant |m \vee [q_m]m''$
$m' \otimes q_m \leqslant m'' \Leftrightarrow m' \wedge m \leqslant [q_m]m''$
$m' \otimes q_m \leqslant m'' \Leftrightarrow (m' \wedge m) \otimes q_m \leqslant m''$
$m' \otimes q_m = (m' \wedge m) \otimes q_m$ ☻

What does this last proposition mean? It says that, starting at m and then there occurring a true statement that m, you get to where you would be if, having taken on board that m is true (moving to $m \wedge m'$), there then occurs a true statement that m. That is certainly the intended effect of the co-precondition, and putting it in that form makes this clearer. Furthermore, there is a connection with PAL. As we remarked at the end of §2.3.3, PAL is a special case of DEL, and hence have algebraic representatives based on their full extensions. Actions then are given a syntax: $\{a_1 \dots a_k\}!\phi$ is the action of announcing to agents $a_1 \dots a_k$ that ϕ is true. $\mathcal{G}!\phi$ is an action α such that $\mu(\alpha) = \neg\phi$ and

$$f_a(\alpha) = \begin{cases} \alpha & if\ a \in \mathcal{G} \\ 1 & otherwise \end{cases}$$

(Recall that 1 is the 'skip' action.) The modality $[\mathcal{G}!\phi]$ is true of its operand precisely when, after a true statement (announcement) that ϕ to the agents in the set $\mathcal{G}$, that operand is true.

Consider the following situation: One agent, a, has no opinion whether p or $\neg p$, and there is one possible announcement, that $p \wedge \neg B_p$, i.e. that p is true, but the agent does not know it. This is representable, as faithfully as the correctness of the epistemic representation thesis of §2.3.1 would seem to allow, as the 1-agent PAL model drawn in diagram 6.2.

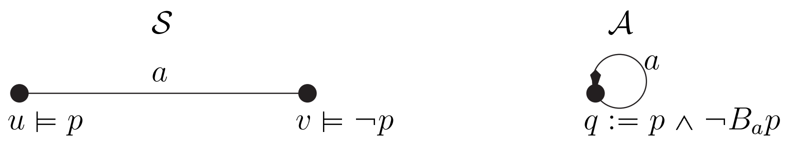


Figure 6.2: A DEL model of the situation in which agent a, unsure whether or not p, learns that $p \wedge \neg B_a p$

That model induces the *full* PAL model (cf. definition 6.2.7) illustrated in figure 6.3.

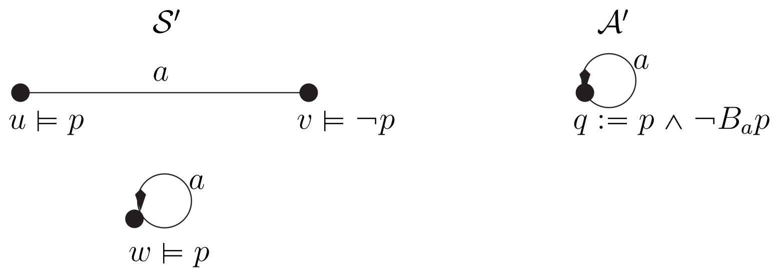


Figure 6.3: The full DEL model induced by the model in figure 6.2

Then by following the recipe given in §6.2.2, we get a 1-agent algebraic expansion model we will call $\mathcal{M}'$, whose lattice of propositions is the 8-element powerset algebra over $\{u, v, w\}$. We will briefly abuse notation here, writing u for $\{u\}$ and so on. Then notice that because $u = (u \vee w) \wedge \neg B_a(u \vee w)$, $q = q_u$. In figure 6.4 we represent the mapping q within $\mathcal{M}'$.

PAL provides a well-studied and precise method for specifying what changes in state a given statement brings about. We want to apply this to the algebraic semantics of which we know PAL to be a special case. However, In PAL, specifying the precondition of an action succeeds in specifying its semantic content, but the same is not true in the more general algebraic semantics. The endomorphism associated with an action by PAL – what we might call its 'operational semantics', which should capture not just the truth-conditions, but the pragmatic effect – depend upon the action's pre-condition, and the update rule (definition 2.3.9). Thus the endomorphism associated with q via the $\otimes$ operator is constrained further, in the representation given in 6.2.2, than just by q's pre-condition. If m satisfies the precondition then it is mapped not to some arbitrary[7] element, but rather to the join of those elements that are q-successors of elements that are stronger than m:

[7] Strictly speaking the elements can never be arbitrary, as the operation must preserve arbitrary joins, but the point remains that there is no flexibility in the DEL case, while there is in the general case.

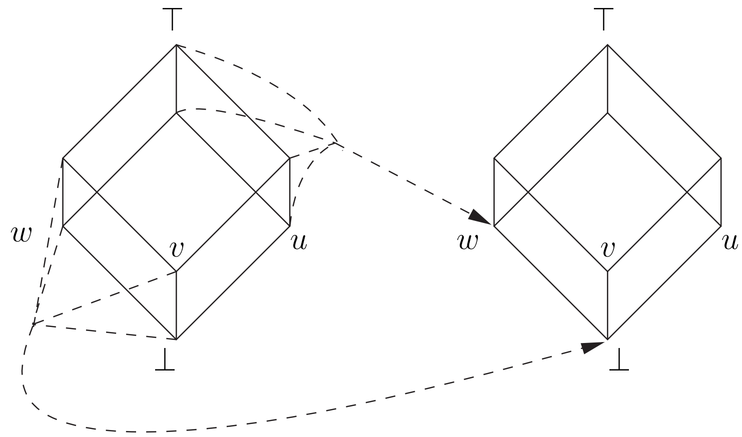


Figure 6.4: The $\otimes q_u$ endomorphism on $\mathcal{M}'$

$$\begin{array}{llll} \otimes q: & M & \to & M \\ & m & \mapsto & \begin{cases} \bot & \text{if } m \in ker(q) \\ \bigvee\{m'^\frown a | a \leqslant q \ \& \ m' \leqslant m\} & \text{otherwise} \end{cases} \end{array}$$

We get this because the rule given for product update (definition 2.3.9) states that if a point w satisfies the pre-condition of an atomic action α, then that action maps that state to $\langle w, \alpha\rangle$; otherwise it is not in the domain of the function. Such a rule is unavailable as things stand in the algebraic setting. We will explore in more detail the algebraic structures that result from lifting DEL models as exemplified above, according to the recipe of §6.2.2 later, in §7.2. We now leave the question of the meaning of actions, moving to a framework in which the co-precondition semantics are adequate to define the meaning of an action: We will consider only (a restricted class of) agent-less algebraic models, in which the pragmatic effect of an action is no longer problematic: Epistemic actions do not change the environment when there is no epistemic agent in the environment.

6.3.3 AGM algebraic models

In a very restricted class of algebraic models, which contain no agents and in which every element of the lattice of propositions is a (ground) fact, we find a close parallel with the AGM postulates.

Definition 6.3.7 (Fact):
An element p of an algebra of propositions is a *fact* in a proposition action system $\langle \mathcal{M}, \mathcal{Q}, \otimes\rangle$ iff$_{df}$ $\forall q \in Q, m \otimes q \leqslant m$.

Definition 6.3.8 (n-agent Factual Algebraic Model):

An n-agent algebraic model is *factual* iff$_{df}$ every element m in its lattice of propositions is a fact.

Consider the following equivalents of the AGM postulates:

(AF1) $m \circledast q_{m'} \leqslant m'$

(AF2) $m \nleqslant \bigvee ker(q) \Rightarrow m \circledast q_{m'} = m \wedge m'$

(AF3) $m' \neq \bot \Rightarrow m \circledast m' \neq \bot$

(AF4) $m' = m'' \Rightarrow (m \circledast q_{m'} = m \circledast q_{m''})$

(AF5) $(m \circledast q_{m'}) \wedge m'' \leqslant m \circledast q_{(m' \wedge m'')}$

(AF6) $(m \circledast q_{m'}) \wedge m'' \neq \bot \Rightarrow m \circledast q_{(m' \wedge m'')} \leqslant (m \circledast q_{m'}) \wedge m''$

The first four of these are respected in an extensional 0-agent factual model:

Definition 6.3.9 (AGM Algebraic Model):
A 0-agent factual algebraic model is an *AGM algebraic model* iff$_{df}$ it is extensional (see definition 6.3.5), and its lattice of propositions is Boolean and atomic.

Proposition 6.3.1:
In an *AGM* algebraic model, (AF1)–(AF4) are all respected (when the corresponding actions q_m exist), but (AF5) and (AF6) can not be respected.

Proof. By hypothesis the model is factual, Boolean, atomic and extensional.

(**AF1**) By clause ($\circledast$1) of definition 6.3.1, we have $\exists p : m \circledast q_{m'} = p \otimes q_{m'}$.
Proposition 6.3.2 gives us $p \otimes q_{m'} \leqslant (p \wedge m') \otimes q_{m'}$,
which, because $\leqslant$ is sup-preserving, yields $m \circledast q_{m'} \leqslant m' \otimes q_{m'}$,
and since by hypothesis m' is factual, $m' \otimes q_{m'} \leqslant m'$,
so indeed $m \circledast q_{m'} \leqslant m'$.

(**AF2**) Suppose $m \nleqslant \neg m'$; then $m \nleqslant \bigvee ker(q_{m'})$.
So $m \otimes q_{m'} \neq \bot$.
So by clause ($\circledast$2) of definition 6.3.1, we have $m \circledast q_{m'} = m \otimes q_{m'}$;
but by proposition 6.3.2, $m \otimes q_{m'} \leqslant (m \wedge m') \otimes q_{m'}$.
Now because $\leqslant$ is sup-preserving, we have $(m \wedge m') \otimes q_{m'} \leqslant m' \otimes q_{m'}$. And since m and m' are factual: $m \otimes q_{m'} \leqslant m$ &
$m' \otimes q_{m'} \leqslant m'$. So $m \circledast q_{m'} \leqslant m$ and $m \circledast q_{m'} \leqslant m'$. I.e. $m \circledast q_{m'} \leqslant m \wedge m'$.

(**AF3**) Assume that $m' \neq \bot$. Then $\neg \bigvee ker(q_{m'}) \neq \bot$, so there is some m'' such that $m'' \otimes q_{m'} \neq \bot$. And so $q_{m'} \neq \bot$.
Then by the contrapositive of ($\circledast$3) from definition 6.3.1, $m \circledast q_{m'} \neq \bot$.

(**AF4**) Assume that $m = m'$. Then clearly $m'' \wedge m = m'' \wedge m'$. But if the lattice is atomic then $p \otimes q_{p'} = p \wedge p'$:

The $\leqslant$ direction is immediate, as the system is factual, so $p \otimes q \leqslant p$, and we have already seen that $p \otimes q_m \leqslant m$.

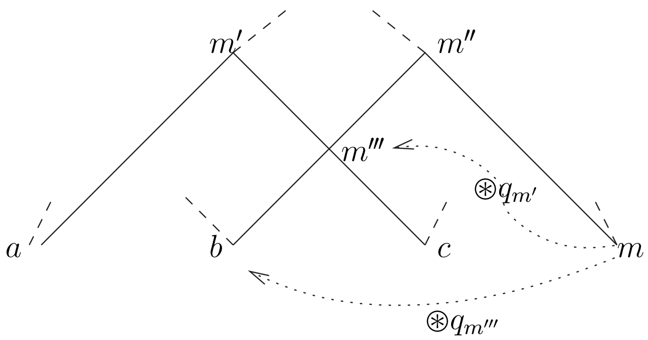
The $\geqslant$ direction uses the atomicity of the lattice:

$$\begin{aligned} p \wedge m &\leqslant \bigvee\{a \in Atm(M) | a \leqslant p \wedge m\} \\ &= \bigvee\{a \in Atm(M) | a \leqslant p \ \& \ a \leqslant m\} \\ &= \bigvee\{a \in Atm(M) | a \leqslant p \ \& \ a \nleqslant \neg m\} \\ &= \bigvee\{a \in Atm(M) | a \leqslant p \ \& \ a \notin ker(q_m)\} \\ &= p \otimes q_m. \end{aligned}$$

Even if the revision operator were to behave well, in the sense that when two actions have the same kernel, they operate in the same manner through it (as is the case in extensional models for example), we would still find that (AF5) and (AF6) do not hold:

(**AF5**) Does not hold when e.g. $m \wedge m' = \perp$, with $m' = c \vee d \vee e$ and $m'' = d \vee e \vee f$; $m \circledast q_{d \vee e} = d$ and $m \circledast q_{m'} = d \vee e$. For example, consider the following partially illustrated example:

Figure 6.5: AGM Algebraic model counter-example to (AF5)



(**AF6**) Does not hold for example in the following case: $m \wedge m' = \perp$, with $m' = a \vee b \vee c$ and $m'' = a \vee c \vee d$, where $a \dots d$ are atoms; $m \circledast q_{m' \wedge m''} = m \circledast q_{c \vee e} = c \vee e$ and $m \circledast q_{c \vee d \vee e} = c$.

☻

6.3.4 Compound Revision

Baltag and Sadrzadeh remark that

> *[the theory acquired after a revision, $m \circledast q$] can be thought of as the result of updating some previously existing situation with the actual experiment q. The situation m' expresses some tentative theory about the original state of the world. This tentative theory is consistent with the result of the new experiment q, that is $m' \otimes q \neq \bot$. In other words, this expresses a possible "static belief revision" of m (with the information provided by q).* ([5], p.7)

The heart of belief revision lies in describing what this 'some tentative theory' is. We can introduce another operator, $\owedge$, which will be the static revision operator, so that

$$(\star 1) \quad m \circledast q = (m \owedge q) \otimes q$$

$$(\star 2) \quad q \circledast^{Q} q' = (q \owedge q') \bullet q'$$

[($\star$1)] can be compared with proposition 6.3.2; notice that in both cases we distinguish between the semantic and pragmatic parts of an action. The semantic part tells us how the world is, and is the first one two be applied; the pragmatic part tells us that the action itself has just occurred.

Now this tentative theory can be either the statically revised *or* the statically contracted theory. Such operators 'pre-encode' (cf. [7], p.11) what would happen when such actions of revision do take place. They are thus also closely related to the $[*-]$ operator of DDL. One might think that abstracting away from the dynamic nature of the revision, hopping on the train of thought from last quotation, would help to make things clearer regarding completeness. However, this is not the case.

We will briefly consider here the problem of finding a complete proof system for the revision operators, as defined by the axioms in definition 6.3.1. In fact we will restrict our attention to the operator on the lattice of propositions; the problems for the two are very closely analogous. The first axiom was

$$(\circledast 1) \quad \exists m' m \circledast a = m' \otimes q$$

If we define $\circledast$ as a compound, as in ($\star$1) above, this is for free. All of the remaining axioms but one are also straightforward. For example if we write the following equivalent (contrapositive) form of ($\circledast$3):

$$(\circledast 3') \quad m \circledast a \leqslant \bot \Rightarrow m \leqslant \bot$$

then we can straightforwardly write it as a proof rule. We could also pass via our definition of $\circledast$ as a compound if we wish:

$$(\circledast 3'') \quad m \owedge a \leqslant [a]\bot \Rightarrow m \leqslant \bot$$

Similarly, ($\circledast$4) is immediately tractable, as

$$\frac{\Gamma \vdash m}{\Gamma \vdash m \oslash a}$$

The difficult part of the proof theory is of course the following axiom:

($\circledast$2) $m \otimes a \neq \bot \Rightarrow m \circledast a \leqslant m \otimes a$

which is implied, given ($\star$1), by

($\circledast$2') $m \otimes a \neq \bot \Rightarrow m \oslash a \leqslant m \wedge \neg[q]\bot$

In ($\circledast$2') we have used Boolean negation, just to force the $\oslash$ operator to be a static *revision* operator in a standard sense. We could equally achieve $\circledast$2 by way of the following version, in which $\oslash$ is a contraction operator:

($\circledast$2") $m \otimes a \neq \bot \Rightarrow m \oslash a \leqslant m$

This is non-trivial because the antecedent has a consistency check, and the consequent a strength claim. This time if we take the contrapositive we get the same situation. We cannot have:

$$\frac{m \nvdash [q]\bot}{m \oslash a \vdash m}$$

(If we were to limit ourselves to a finite number of proofs to be searched, this would be legitimate, but then it would not be very interesting.) If we could specify rules for *atoms* then we could immediately have:

$$\frac{a \vdash m \otimes a}{m \oslash a \vdash m}$$

It is precisely this sort of thing which we can do in a modal logic, via the $\Diamond$ operators. We would like to say that if $w \models \langle q\rangle\top$ then $w \vdash (m \otimes a) \supset m$; i.e. that

$$\langle q\rangle\top \supset (m \otimes a \supset m)$$

6.3.5 Concluding Remarks

We have summarised and discussed this algebraic approach to dynamic belief revision. We considered ways to interface the algebraic approach more directly the AGM tradition, with DDL and preference logic. We have also illustrated the beginnings of a way in which an operational semantics for actions can be given in algebraic terms. That consideration of semantics led us to suggest that the axiomatisation of the revision operator as it occurs in [5] does not fully respect the AGM postulates, more specifically does not respect the last two. It is of interest to compare the discussion of the semantics of actions with the contents of [24], though I wonder if all of us have not being abusing the word 'semantics'. It strikes me that it has been, at least in part, *pragmatics* that were our real concern, and the concern on the horizon of [24]. Although we leave behind considerations of meanings of actions, we turn now to look at the kind of structures presented in [24], in which *time*, or some ordering over events, plays a more explicit role than it has done in any of the mathematical structures we have so far seen.

7 Message-Passing: Trust, Confidence and Time

7.1 Introduction

If revisions are actions then like other actions they take place over time, and in a temporal position, relative to other things. There is a sense in which the Moore problems emphasise the first of those consequences: Actions take place over time, so at the end of an action the world is not as it was before the action. In this chapter we consider the second consequence: Actions are different depending on when they take place, for example relative to other actions.

7.2 Direct Temporal Reasoning

7.2.1 Why Time

In DEL and in the algebraic generalisation of it, we find good ways to treat the semantics of events, and in the latter, with its revision operator, some limited facility for modeling pragmatic reasoning about events: Agents can consider that they were wrong about the state that they thought they were in, given the semantics of the incoming event, and this sometimes represents the agent thinking that she was mistaken about a given event. But as we mentioned in §2.2.3, we want to reason more fully about pragmatic aspects of epistemic actions, in particular faithfully representing actions not just as anonymous statements, but as statements from other agents in the system. Consider the following statement:

> *I am becoming convinced that 'belief' may not be the right notion in understanding our dynamic handling of incoming information. Perhaps [...] incoming signals 'in favour of P' do not immediately make us believe that P, but rather add* grounds for believing*, which may eventually add up to true belief.* ([7], p. 23)

In a real situation of communication between humans, that description is more appropriate than the abstract description provided by DEL and modelings in that spirit. This is relevant to the establishment of a model of the kind that would fulfill the wishes of a doxastic version of Baltag and Moss' representation thesis (§2.3.2). If I am disinclined to believe a proposition p, and you, whom I barely trust, tell me that p, then it is unlikely that I will believe that p. But I will know that you told me that p, and presumably if enough people tell me that p then, unless I am remarkably stubborn, or have some good grounds for believing $\neg p$, then I should believe that p. Is it enough, then, to slowly upgrade the p points with these statements? Such an approach could certainly be taken in the rope model framework. However, suppose now that I do trust you, and you tell me that p, and I believe it. But suppose then that I learn, from some source whom I consider even more trustworthy than you, that you are a notorious liar. Then I will want not to believe that p any more, or rather: I will not want to believe *p only because you have told me that p*. But if all that I had done was to eliminate the $\neg p$ points (PAL; §2.3.2) or upgrade the p points (soft change; §4.3.3) in my belief state, then there would be no way for me to do this, as I would have no record of what I had upgraded and why. Perhaps that does reflect part of the psychological reality: We are not so rational, or perhaps so computationally capable, and with such perfect recall, that we can suddenly 'disbelieve' all and only the statements that come (only) from a very unreliable source. However, we do want a logical model in which such an activity can be represented.

One way to do this is to make agents reason explicitly about time and events. One aim of this section is to show that in the algebraic setting presented in §6 it is possible to represent such reasoning. We will consider a restricted class of the structures introduced in that section, temporal proposition-action systems:

Definition 7.2.1 (Temporal Proposition-Action System):
A propositional-action system $\langle \mathcal{M}, \mathcal{Q}, \otimes \rangle$ is *temporal* iff $\forall a, a' \in Atm(M), \forall q \in Q, a \neq a' \Rightarrow (a \otimes q \neq \bot \Rightarrow a \otimes q \neq a' \otimes q)$.

Remark 7.2.1:
The proposition-action systems underlying the algebraic models generated in the representation theorem in section 6.2.2 are temporal.

The representation theorem for dynamic epistemic logic, in section 6.2.2, is rather heavy-handed, in the sense that in the general case, given some DEL model $\langle \mathfrak{M}, \mathfrak{A} \rangle$, the algebraic representative contains a great number of elements, both in the lattice of propositions $\mathcal{M}$ that have no interesting correlate in $\mathfrak{M}$. They have no interesting correlate in the sense that they can never be the case, nor thought by any of the agents to be the case. They correspond to unions of points from different 'moments in time'. An example would be the union of a point $s \in \mathfrak{M}$ with a point $\langle s, a \rangle \in \mathfrak{M} \boxtimes \mathfrak{A}$, where (as is usually the case) $\langle s, a \rangle \notin \mathfrak{M}$ and $s \notin \mathfrak{M} \boxtimes \mathfrak{A}$.

There is no 'eternal return': each atom specifies its history. This suggests that we can represent a different sort of action semantics, the 'history-based' epistemic temporal logic of [24], using these algebras.

We will find that *branching time* versions of the history-based structures of [24] correspond in some sense to full DEL models. (For a more serious examination of the relation between branching and linear time structures and their logics, see for example [15].)

We first present the *linear time* semantics given in [24] (7.2.2), then the slightly different branching time structures in which we will be interested and how they relate to the former (7.2.3); then we illustrate how these models are related to full DEL models (§7.2.4; cf. §6.2.2).

7.2.2 History-based models

We start by taking any non-empty set of atomic and unanalysed events Σ. An infinite 'history' is just an unending sequence of events:

Definition 7.2.2 (Infinite History):
H is an *infinite history* over Σ iff$_{df}$ $H \in \Sigma^\omega$.

We use the following notation: Where $k \in \omega$, $H_k = H \restriction (k+1)$. Unsurprisingly, we call such finite sequences 'finite histories'.

Definition 7.2.3 (Finite History):
A *finite history* of length k is a sequence $\langle \sigma_1 \dots \sigma_k \rangle \in \Sigma^k$. The set of all finite histories is denoted Σ^*. We write the concatenation of two finite histories $h = \langle \sigma_1 \dots \sigma_m \rangle$ and $h' \langle \sigma'_1 \dots \sigma'_n \rangle$ as $h \frown h'$, $= \langle sigma_1 \dots \sigma_m \sigma'_1 \dots \sigma'_n \rangle$. We abbreviate $\langle e \rangle$ to e. The empty history is denoted 1.

Given such a set of events Σ, we also specify a *protocol* $\subseteq \Sigma^*$, which can be thought of as the *legal* sequences of events. For example, a game of chess might be analysed by saying that there is a set of events Σ_C, which correspond to the moves permitted: for example, moving a pawn two steps forward; but certain sequences of events (e.g. two white pieces moving in succession) might be disallowed by the rules, and so would not, in this analysis, be part of the protocol. It is of course a metaphysical question whether we might insist that a protocol applies outside the realm of games, in the real world:

Definition 7.2.4 (Binary Protocol):
$\mathbb{P}$ is a *binary protocol* over Σ iff$_{df}$ $\mathbb{P} \subseteq \Sigma^\omega$.

An idea behind the history-based structures of [24] is that for each agent some events are 'visible' and others 'invisible'. So they define what we will call 'infinite history systems' as follows:

Definition 7.2.5 (n-agent History System):

$\mathcal{S} = \langle \mathbb{P}, \{E_a\}_{a \in n}, V \rangle$ is an n-agent history system over Σ iff$_{df}$ $\mathbb{P}$ is a binary protocol over Σ and $\forall a \in n, E_a \subseteq \Sigma$.

Here each E_a represents the events of which agent a can be, and always is, aware of the occurrence. Each E_a then induces a local view function between finite histories; we first define inductively the projection function of an agent on the basis of E_a. The projection of h for a is just the events of h that a can see, in the order in which they occur in h:

$\lambda_a(1) = 1$

$$\lambda_a(h^\frown e) = \lambda_a(h)^\frown \begin{cases} e & if\ e \in E_a \\ 1 & otherwise \end{cases}$$

[24] calls the range of a's projection function a's 'local histories'. There is a sense in which a's local history at h (i.e. $\lambda_a(h)$) represents the way in which agent a sees h: She has seen those events of which she can be aware, and is entirely ignorant of those of which she cannot be aware. We say '*entirely* ignorant' because it is not the case that a can be correctly characterised by saying that she thinks that some events have not taken place. Rather, those events are, as it were, beyond her ken: She is unable to entertain the possibility of them having occurred. There is a sense in which they are not part of her environment. This should become clearer when we define a's local view function on the basis of the projection function. That will be a function between finite histories allowed by the protocol and sets of such finite histories. It will be useful to have some additional notation: We define the relation $\rightsquigarrow \subseteq 2^{\Sigma^*} \times 2^{\Sigma^\omega}$:

Definition 7.2.6 (Finite Prefix):
For a finite history h and a (finite or infinite) history H, we say that h is a finite prefix of H, written $h \rightsquigarrow H$, iff$_{df}$ $\exists H' : H = h^\frown H'$.

Then the function $FinPre : 2^{\Sigma^* \cup \Sigma^\omega} \rightarrow 2^{\Sigma^*}$, giving the set of finite prefixes of a set of histories:

Definition 7.2.7 (Finite Prefixes):
$FinPre(\cup_i H_i) = \cup_i \{h | \exists H : h^\frown H = H_i\}$

Now we can define the local view function induced by a projection function:

$$\begin{array}{rccc} f_a(h): & FinPre(\mathbb{P}) & \rightarrow & 2^{FinPre(\mathbb{P})} \\ & h & \mapsto & \{h' | \lambda_a(h) = \lambda_a(h')\} \end{array}$$

That is, the points (histories) which agent a does not rule out are precisely those that appear the same to her as the actual history. Thus the agent can never rule out the actual history, which will always appear to her as it appears to her. This is what we meant by saying that the agent is entirely ignorant of those events of which she is not aware, and it entails that such events do not lead her to false beliefs simply because she is not aware of their occurrence.

Definition 7.2.8 (n-agent History Structure):
$\mathcal{S} = \langle \mathbb{P}, \{f_a\}_{a \in n} \rangle$ is an *n-agent history structure* over Σ iff$_{df}$ $\mathbb{P}$ is a binary protocol over Σ $\forall a \in n, f_a : FinPre(\mathbb{P}) \to 2^{FinPre(\mathbb{P})}$.

Clearly an n-agent history system defines uniquely an n-agent history structure. Furthermore, given an n-agent history structure defined by some n-agent history system, that n-agent history system is uniquely determined. Notice that if we set things up differently, allowing ourselves to define directly the λ_a function (as is done in [23]), rather than via the E_i functions, then we would not be able to retrieve that λ_i function if we were given only the induced local view function. All that is important from the point of view of a, or at least from the point of view of her epistemic state, is that she has some way of distinguishing between histories, even if she 'sees' them as different from the way they are. All that she needs in order to be omniscient, then, is for λ_i to be bijective. We will later allow the f_a functions to be defined independently, without reference to any E_i function or to λ_i's; before that, we look at the way these structures are used as the models for a formal language. In order to make the structures models for a language whose propositional variables are Φ, they must be supplemented with a valuation function:

Definition 7.2.9 (n-agent History Model):
$\langle \mathbb{P}, \{f_a\}_{a \in n}, V \rangle$ is an *n-agent history model* over Σ iff$_{df}$ $\langle \mathbb{P}, \{f_a\}_{a \in n} \rangle$ is an n-agent history structure over Σ and $V : FinPre(\mathbb{P}) \to \Phi$.

[24] introduce two temporal modal operators into their language $\mathcal{L}_T$: A unary operator O, meaning 'at the next instant', and a binary operator U, where $\phi U \psi$ is to be interpreted as ϕ holds until ψ is true. This necessitates giving the semantics of that language with respect not to finite histories but to infinite histories indexed by an ordinal giving the position in that history, i.e. tuples $\langle H, k \rangle$. This is because for example the finite history h does not contain enough information to express the intended meaning of Op because its truth depends on what the next state is, which is not determined by h. This expressive power is however never used in [24], so after stating the semantics provided there, we will follow up and make formal a remark made by the authors to the effect that one can evaluate most such formulae with respect to finite histories rather than infinite histories supplemented by ordinals. Again, we suppress the subscript $\mathfrak{M}$ in $\langle H, k \rangle \models_{\mathfrak{M}} \phi$. The first four clauses in the following truth-condition definition are essentially that same as those given in 2.3.1.

1. $\langle H, k \rangle \models p \Leftrightarrow p \in V(H_k)$
2. $\langle H, k \rangle \models \neg \phi \Leftrightarrow \langle H, k \rangle \not\models \phi$
3. $\langle H, k \rangle \models \phi \wedge \psi \Leftrightarrow \langle H, k \rangle \models \phi \ \& \ \langle H, k \rangle \models \psi$

4. $\langle H, k\rangle \models B_a\phi \Leftrightarrow \forall\langle H', k'\rangle : H'_{k'} \in f_a(H_k), \langle H', k'\rangle \models \phi$

5. $\langle H, k\rangle \models O\phi \Leftrightarrow \langle H, k+1\rangle \models \phi$

6. $\langle H, k\rangle \models \phi U\psi \Leftrightarrow \exists m \geqslant k \forall l \geqslant k(l < m \Rightarrow (\langle H, l\rangle \models \phi \wedge \langle H, m\rangle \models \psi))$

7.2.3 The locally temporal fragment

Parikh and Ramanujam write that they 'occasionally abuse language and write $h \models B_a\phi$ when [they] mean $\langle H, k\rangle \models B_a\phi$' (p. 457). Presumably they mean H_k instead of h, but apart from that (presumably unintentional) abuse, what they mean to do is to be tolerated. They point out that in order to assess whether a formula $B_a\phi$ is true, one need only know the finite history at which it is being evaluated. Notice that this is also true for atomic formulae p. Then the only formulae for which it is not true are those in which there occurs a temporal operator that is not 'shielded' by an epistemic operator. No such formulae occur explicitly in [24]. Let us call the formulae in which all temporal operators do occur within the scope of an epistemic operator 'locally temporal', because they are only temporal with respect to local histories. We provide an obvious alternative semantics for all of the formulae of $\mathcal{L}_T$, which for the locally temporal fragment is equivalent to the semantics given above. In doing this, we also show the connection between locally temporal models and relational models: The former are a special case of the latter, supplemented by a temporal accessibility relation.

Now we observe that all of the non-temporal information we can get from an n-agent history structure can be contained by a relational model over the set of finite prefixes of the system's protocol. Specifically, the n-agent relational model generated by the n-agent history model $\langle \mathbb{P}, \{f_a\}_{a\in n}, V\rangle$ is the n-agent relational model $\langle FinPre(\mathbb{P}), \{f_a\}_{a\in n}, V\rangle$.

The difference between an epistemic temporal relational frame (model) and a standard relational frame (model) is that in the former the set of points has the structure of a history.

Definition 7.2.10 (n-agent Epistemic Temporal Relational Model):
$\langle W, \{f_a\}_{a\in n}, \leqslant, V\rangle$ is an *n-agent epistemic temporal relational model* iff$_{df}$ $\langle W, \{f_a\}_{a\in n}, V\rangle$ is an n-agent relational model, and $\leqslant \subseteq W \times W$ is a partial order (reflexive, transitive and anti-symmetric).

The n-agent epistemic temporal relational model generated by the n-agent history model $\langle \mathbb{P}, \{f_a\}_{a\in n}, V\rangle$ is just $\langle FinPre(\mathbb{P}), \{f_a\}_{a\in n}, \rightsquigarrow, V\rangle$.

The definition of truth for $\mathcal{L}_T$-formula in an n-agent epistemic temporal relational model includes that for relational models, and supplements it with the following two clauses:

(5') $h \models O\phi \Leftrightarrow \forall h'((h \rightsquigarrow h' \;\&\; \forall h'' \neq h'(h'' \rightsquigarrow h' \Rightarrow h'' \rightsquigarrow h)) \Rightarrow h \models \phi)$

(6') $h \models \phi U \psi \Leftrightarrow \exists h' : h \rightsquigarrow h' \;\&\; \forall h'' \neq h'(h \rightsquigarrow h'' \rightsquigarrow h' \Rightarrow (h'' \models \phi \;\&\; h' \models \psi))$.

These make the temporal modalities into branching time modalities, where in the original definition of [24] they were linear time modalities . The point that we make here is that they are equivalent when we only consider them as occurring within the scope of the epistemic modalities:

Remark 7.2.2:
For any ϕ in the locally temporal fragment of $\mathcal{L}_T$, $\langle H, k\rangle \models \phi$ in an n-agent history model $\mathcal{M}$ iff $H_k \models \phi$ in the n-agent epistemic temporal relational model generated by $\mathcal{M}$.

Therefore a branching time semantics is at least as appropriate as a linear time semantics as far as the locally temporal fragment is concerned. And it is the locally temporal fragment which is of interest when modeling agents' own reasoning about time, which is what is of interest here.

7.2.4 Algebraic Temporal Operator

[6] provides a representation theorem for DEL models as history models. Here we make the simple observation that the particular kind of full DEL models that are used to prove the representation theorem of [26] have in them enough structure to provide a branching-time semantics. To see this, simply take the ordering given by concatenation.

This emphasises that what we qualified as the 'temporal' nature of a class of algebraic structures (definition 7.2.1) is appropriately named: It is such that the monoidal operator acts as concatenation in the natural sense. It is then possible to define standard temporal operators in terms of it. For example:

$$F(m) = \bigvee\{m' | \forall q m' \otimes q \leqslant m\}$$

To see that this is correct, notice that it is equivalent to $F(m) = \bigvee\{m' | \forall q \forall m'' m' \otimes q = m'' \Rightarrow m'' \leqslant m\}$, so that $m' \leqslant F(m) \Leftrightarrow \forall q \forall m''(m' \otimes q = m'' \Rightarrow m'' \leqslant m)$, which should recognisably equivalent to the truth condition of the branching time future operator. We can give similar definitions for algebraic versions of the O and U operators, cf. (5') and (6') above.

7.3 Indirect Temporal Reasoning

Adding temporal reasoning increases complexity[1], but we do not need so much in order to capture the kind of reasoning involved here. Rather than store the whole history in a model, we can have (propositional) variables that in effect store some parts of the history. We will naturally want to choose to label those parts which are considered 'important', chopping things up in a rational way. Here we suggest one such natural way to do that. If we make a number of simplifications about consistency of behaviour, we can show that a partial form of the epistemic representation thesis is satisfied by a version of the public announcement preference logic of §4.3.3.

What exactly is the 'hard information' that can be acquired from a statement by you that some proposition, ϕ say, holds? I would propose the following: If you are honest, then you believe that ϕ. So if we make the simplifying assumption that your honesty is invariant over time, then we can represent it in a PAPL model by a propositional variable, H_{you}. Then we can model the 'internal announcement' of ϕ by j as an elimination of all those points in which $B_j\phi$ does not hold, and in which H_j does hold. That is, j's announcement to a group $\mathcal{G}$ that ϕ just is the public announcement action represented by the modality $[\mathcal{G}!(H_j \supset \phi)]$.

I propose that this modeling, in the context of public announcement preference logic, is appropriate enough for a restricted class of simple situations as to satisfy a doxastic version of the Baltag and Moss' thesis. Of course there are also very many less simple situations. Part of the tenacity of the original representation thesis is the immense diversity of possible epistemic situations and actions. The situation is no cleaner in the case of beliefs and hence of the doxastic version. Some more involved or specific approach will probably suffice in each case to satisfy the representation thesis (in its current, unformulated limbo status). However, ideally one would have a mathematically precise class of models cited in the doxastic representation thesis, just as Baltag and Moss are able to in theirs. Chapters 3 and 4 ended with a suggestion for future research in precisely this area, and part of the force of this section is the same.

[1]This remark is made more precisely in [6].

8 Conclusion

Much of this thesis has been taken up by presenting different pieces of recent work that have been done in the field of belief revision and adapting them to a multi-agent setting. It shows that dialogue is possible between traditions as different as those behind, for example, Shoham and Maynard-Reid II's work on representing propositional belief fusion; Baltag Coecke and Sadrzadeh's work on algebraic epistemic logic; and Lewis' work on logical analysis of counterfactual conditionals.

Exposition and integration is not all that we have done however. A running theme of the thesis has been the importance of modeling actions correctly. DDL does not do this as one might expect it to; perhaps no surprise, as it is built on the AGM postulates. But it is possible to respect the spirit of the AGM postulates and still have a satisfactory representation of actions. This can be done, as ($\star$1) in §6.3.4 suggests, and as the modelings of soft information change of §4.3.3 testify more confidently.

We also began sketching an expressive logic for representing what we called, following Shoham and Maynard-Reid II, 'anonymous belief states', and showed that it has the capability to express (some forms of) *merging*.

A significant part of the work of the thesis was in working out the completeness proof of Appendix 5. This turns out to be tangential to much of the rest of the thesis, but could have made the centre-piece of another thesis. The reduction axioms for public announcement proposed in §4.3.3 lead directly for a completeness proof for an extended logic, one which can reason about simple model-changing operations.

Most of the chapters have left questions primed for future research. In chapter 3 we raised the issue of finding a mathematical structure amenable to logical analysis and specifically suited for representing the different possible sorts of merging or fusion of pre-orders. A related issue was left open at the end of chapter 4: What kind of logic should we use for expressing preference changes, in the full generality in which they can occur? Chapter 6 left two very different issues open: First was the sequent rules for the revision operator, or perhaps a more restricted operator from which it might be definable. Second was the rather more vague question about the meaning of actions in the algebraic setting, which requires more conceptual clarity before it can be pursued. Finally in §7 we found an object for a restricted doxastic version of Baltag and Moss' epistemic representation thesis. Formulating and justifying some doxastic version of that thesis is something that could very legitimately be pursued by any researcher in the field of these kinds of logics who is at all concerned with their relevance to the representation or modeling of social interactions.

Bibliography

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson, *On the logic of theory change: Partial meet contraction and revision functions*, Journal of Symbolic Logic **50** (1985), no. 2, 510–530.

[2] Alexandru Baltag, Bob Coecke, and Mehrnoosh Sadrzadeh, *Epistemic actions as resources*, Journal of Logic and Computation (2006), to appear.

[3] Alexandru Baltag and Lawrence S. Moss, *Logics for epistemic programs*, Synthese **139** (2004), 165–224.

[4] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki, *The logic of public announcements, common knowledge and private suspicions*, Tech. Report SEN-R9922, CWI, 1999.

[5] Alexandru Baltag and Mehrnoosh Sadrzadeh, *The algebra of multi-agent dynamic belief revision*, Electronic Notes in Theorerical Computer Science **157** (2006), no. 4, 37–56.

[6] Johan van Benthem and Eric Pacuit, *The forest of common knowledge in action: Towards a common perspective*, 2006.

[7] Johan van Benthem, *Dynamic logic for belief change*, manuscript, 2006.

[8] Johan van Benthem, Sieuwert van Otterloo, and Olivier Roy, *Preference logic, conditionals, and solution concepts in games*, **PP-2005** (2005), no. 28.

[9] Patrick Blackburn, Maarten de Rÿke, and Yde Venema, *Modal logic*, Cambridge University Press, 2001.

[10] Adnan Darwiche and Judea Pearl, *On the logic of iterated belief revision*, Artificial Intelligence **89** (1997), no. 1–2, 1–29.

[11] Hans van Ditmarsch and Barteld Kooi, *The secret of my success*, Synthese **151** (2006), no. 2, 201–232.

[12] Ronald Fagin, Joseph Halpern, Moshe Vardi, and Yoram Moses, *Reasoning about knowledge*, MIT Press, Cambridge, MA, USA, 1995.

[13] Eduardo L. Fermé and Sven Ove Hansson, *Selective revision*, Studia Logica **63** (1999), no. 3, 331–342.

[14] Jelle D. Gerbrandy and Willem Groeneveld, *Reasoning about information change*, Journal of Logic, Language, and Information **6** (1997), 147–169.

[15] Valentin Goranko and Alberto Zanardo, *From linear to branching-time temporal logics: Transfer of semantics and definability*, 2006.

[16] Adam Grove, *Two modellings for theory change*, Journal of Philosophical Logic **17** (1988), no. 2, 157–170.

[17] Jaakko Hintikka, *Knowledge and belief*, Cornell University Press, 1962.

[18] Hirofumi Katsuno and Alberto O. Mendelzon, *Propositional knowledge base revision and minimal change*, Artificial Intelligence **52** (1991), 263–294.

[19] Dexter Kozen and Rohit Parikh, *An elementary proof of the completness of pdl.*, Theoretical Computer Science **14** (1981), 113–118.

[20] David Lewis, *Counterfactuals*, Blackwell, 1973.

[21] Pedrito Maynard-Reid II, *Pedigreed belief change*, Ph.D. thesis, Stanford, 2001.

[22] Pedrito Maynard-Reid II and Yoav Shoham, *Belief fusion: Aggregating pedigreed belief states*, Journal of Logic, Language, and Information **10** (2001), no. 2, 183–209.

[23] Eric Pacuit, Rohit Parikh, and Eva Cogan, *The logic of knowledge based obligation*, to appear in *Knowledge, Rationality and Action*, 2006.

[24] Rohit Parikh and Ramaswamy Ramanujam, *A knowledge based semantics of messages*, Journal of Logic, Language, and Information **12** (2003), no. 4, 453–467.

[25] Hans Rott, *Shifting priorities: Simple representations for twenty-seven iterated theory change operators*, Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg (Henrik Lagerlund, Sten Lindstr, and Rysiek Sliwinski, eds.), Uppsala Philosophical Studies, vol. 53, Uppsala, 2006, pp. 359–384.

[26] Mehrnoosh Sadrzadeh, *Actions and resources in epistemic logic*, Ph.D. thesis, Université du Québec à Montreal, 2004.

[27] Krister Segerberg, *Iterated belief revision in dynamic doxastic logic.*

[28] ———, *An essay in classical modal logic*, Filosofiska Studier, vol. 13, Uppsala University, 1971.

[29] ———, *A completeness theorem in the modal logic of programs*, Universal Algebra and Applications (Tadeusz Traczyk, ed.), Banach Centre Publications, vol. IX, PWN (available online at http://journals.impan.gov.pl/bc/), 1982, pp. 31–46.

[30] ———, *Notes on conditional logic*, Studia Logica **48** (1989), no. 2, 157–168.

[31] ———, *A completeness proof in full ddl*, Logic and Logical Philosophy **9** (2001), 77–90.

[32] Wolfgang Spohn, *Ordinal conditional functions: A dynamic theory of epistemic states*, Causation in Decision, Belief Change, and Statistics (W.L. Harper and B. Skyrms, eds.), vol. II, Kluwer, 1988, pp. 105–134.