How to keep on changing your mind, dynamically

Jonathan A. Zvesper

ABSTRACT. Bonanno [Bon07a] presents a modal language $\mathcal{L}_B$ for reasoning about information and changes of belief with a branching next-time semantics. We provide here a complete axiomatisation (theorem 2) of the resulting logic. Bonanno also presents $\mathcal{L}_B$ axioms for the AGM [AGM85] theory change postulates. The AGM postulates say little about *iterated* change. And they are '*static*' in van Benthem's [Ben06] terminology, so not suited for representing the beliefs of introspective single agents or interacting groups of agents. The current paper develops Bonanno's logic to transcend these two features of the AGM postulates. Firstly we give $\mathcal{L}_B$ axioms for various iterated belief revision constraints discussed by Boutilier [Bou96], Darwiche and Pearl [DP97], Nayak et al [NPP03] and others (propositions 5–14). This was explicitly left open in [Bon07a]. Secondly we propose alternative axioms that exploit the temporal vocabulary of the logic to capture the 'dynamic' aspect of revision that is present in van Benthem's [Ben06] proposal.

Keywords: Iterated belief revision, modal logic, branching time, dynamic belief revision

Introduction

A theory is *revised* by a piece of information when that new piece of information is integrated into the theory to form a new theory which is consistent and which accommodates the new piece of information. Alchourrón Gärdenfors and Makinson [AGM85] give what are now know as the 'AGM' postulates for rational revision of a theory. They tell you how to change your mind given incoming information. But they say little about how to *keep on* changing your mind as you receive *successive* pieces of information: they say little about *iterated* belief revision.

Researchers principally from the artificial intelligence community have discussed iterated belief revision, and proposals have been made [Bou96, DP97, NPP03]. Rott [Rot06] provides examples of twenty-seven different methods for iterating belief revision. de Rÿke, van Linder et al [LHM95], and Segerberg [Seg95] have provided object languages for the postulates using modal logic. Some work has been done to connect such proposals with existing logics [Ben06, Lin95, Seg, Zve07].

Bonanno [Bon07a] presents another object language, $\mathcal{L}_B$. It is also modal, but with branching next-time operators and a novel 'information modality' I, where $I\phi$ means that the information received at the current state and time is ϕ. [Bon07a] connects his semantics with the AGM [AGM85] postulates, and explicitly leaves as an open avenue of research the possibility of capturing some of the existing proposals for *iterated* belief revision: 'an interesting topic for future research is whether the principles for iterated revision that have been proposed in the literature can be translated into syntactic axioms.' ([Bon07a], p. 159) In the present paper we explore that avenue, demonstrating extensions of the logic by axioms that capture some constraints on iterated revision. Specifically, we will give syntactic axioms for the principle of 'recalcitrance' [NPP03], the Darwiche-Pearl [DP97] principles, and the lexicographic [NPP03] and 'absolute minimisation' [Bou96, DP97] policies[1]. In addition to those constraints, which are familiar from the literature, we illustrate a policy which we call 'cynicism'.

The AGM postulates were proposed to regulate *theory change*, and not specifically *belief revision*, with its agentive emphasis. Agents have beliefs not just about non-changing states of the world, but also about their own (changing) beliefs, and perhaps even those of other agents. Ann's beliefs after revising by 'it is raining outside and nobody believes it is' should *not* include the belief that nobody believes it's raining, but this is what the AGM postulates would dictate.

Because Bonanno [Bon07a] is concerned with AGM postulates, he restricts revision to pure Boolean formulae, avoiding such problems, which after G. E. Moore we will call 'Moore problems' (cf. [Seg06]). We will consider lifting that restriction, making a proposal for doing what van Benthem [Ben06] and Baltag et al [BS06, BS07] call '*dynamic*' belief revision[2] by remarking that the meaning of an utterance in general concerns the world as it was *before* the utterance was made.[3] We cash out this remark by exploiting the temporal vocabulary of $\mathcal{L}_B$. To give a specific example, we reinterpret the success postulate (($\circledast$1) in section 1.1 below) as '*After* revising by ϕ, the agent believe*s* that ϕ *was* the case,' all of which is expressible

[1] See section 1.1 for the distinction between principle and policy.

[2] These authors use the word 'dynamic' to indicate that epistemic actions like observation or announcement *change* the world. Note that this is not the same as Nayak's use [NPP03] of the word 'dynamic' in a belief revision context.

[3] This remark is not new, indeed it is immediate from the meaning of 'dynamic' in van Benthem's and Baltag and Smets' sense. It also echoes [Yap06].

in $\mathcal{L}_B$. Public announcement logic [Pla89, GG97] is known to handle Moore problems. Therefore in order to demonstrate the validity of our reinterpretation, we will show how to mimic public announcement logic within Bonanno's branching time semantics.

That discussion of dynamic revision takes place below in section 3. Our other main contribution, axiomatic characterisations of several forms of iterated revision, is given in section 2. First we must recall some specific details from the AGM tradition and from Bonanno's work that we will need.

1 Background: AGM and $\mathbb{L}_{AGM}$

1.1 AGM

We assume some underlying formal language $\mathcal{L}$, and present Darwiche and Pearl's [DP97] re-writing of the original AGM postulates, in terms of belief *states*, rather than belief *sets*. The only difference, mathematically speaking, between the two presentations is that in ours we do not identify two belief states just because they have exactly the same $\mathcal{L}$-formulae.[4] That is, the disposition to change one's beliefs, which on this analysis is not part of one's beliefs, is not determined purely by one's belief set. A ***belief set*** is just a set of formulae of some formal language; a ***belief state*** Γ is a primitive entity associated with which there is a belief set $\overline{\Gamma}$.

A ***revision function*** is any function that takes a belief state and a formula and returns a new belief state. Given an entailment relation $\vdash$ over $\mathcal{L}$, the AGM postulates are the following constraints on revision functions:[5]:

$$\Gamma \dotplus \phi \vdash \phi \tag{$\dotplus$1}$$

$$\Gamma \dotplus \phi \subseteq (\Gamma \cup \phi)^{\vdash} \tag{$\dotplus$2}$$

$$\text{If } \Gamma \nvdash \neg\phi \text{ then } (\Gamma \cup \phi)^{\vdash} \subseteq \Gamma \dotplus \phi \tag{$\dotplus$3}$$

$$\text{If } \nvdash \neg\phi \text{ then } \Gamma \dotplus \phi \nvdash \bot \tag{$\dotplus$4}$$

$$\text{If } \vdash \phi \equiv \psi \text{ then } \Gamma \dotplus \phi \equiv \Gamma \dotplus \psi \tag{$\dotplus$5}$$

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

$$\text{If } (\Gamma \dotplus \phi) \nvdash \neg\psi \text{ then } ((\Gamma \dotplus \phi) \cup \{\psi\})^{\vdash} \subseteq \Gamma \dotplus (\phi \wedge \psi) \tag{$\dotplus$7}$$

The language usually considered in the artificial intelligence literature is propositional logic, and the entailment relation classical entailment. In this natural context the AGM postulates are certainly not concerned with

---

[4]Thus although Bonanno uses the belief set version of the postulates, similar results to his are obtainable with respect to the state version.

[5]For ϕ and ψ $\mathcal{L}$-formulae. We abuse notation and write Γ instead of $\overline{\Gamma}$ when no ambiguity arises. We write $\Gamma \equiv \Gamma'$ to mean that $\overline{\Gamma} = \overline{\Gamma'}$.

modelling the introspective or multi-agent aspects of agents. Nonetheless, "Logic is not just about single-agent notions like reasoning, or zero-agent notions like truth, but also about communication between two or more people" (van Benthem [Ben03]). This holds particularly in the analysis of games, and another branch of artificial intelligence, multi-agent systems, calls for development beyond the AGM paradigm. Hence the need to recognise the *dynamic* nature of the interaction between information and beliefs [BS06, BS07, Ben06].

In the case of a single non-introspective agent, the AGM postulates constrain ways to revise a theory given a 'one-shot' notion of revision normatively: revisions which violate the postulates are deemed to be irrational. However, the AGM postulates say little about iterated revision:

REMARK 1. Nayak et al [NPP03] explicitly give the 'only interesting inference about iterated belief change that we can draw from the AGM postulates', viz. $\Gamma \dotplus \phi \nvdash \neg\psi \Rightarrow (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus (\phi \wedge \psi)$ (op. cit., p. 196).

As we have mentioned in the introduction, researchers have discussed various constraints that can be placed on iterated revision, giving postulates that include expressions of the form $(\Gamma \dotplus \phi) \dotplus \psi$, or specifying semantic constraints on iterated revision. Some constraints determine uniquely the iterative version of the revision function, given the one-shot version of the revision function We will call such constraints 'policies'. For example, the lexicographic constraints given by Nayak et al [NPP03] describe a policy (see section 2.3).

The AGM constraints on *one-shot*, *static* revision are given axiomatic expression in a temporal modal logic by Bonanno [Bon07a].

1.2 $\mathbb{L}_{AGM}$

The modal language $\mathcal{L}_B$ has, in addition to a set Φ of proposition letters and the standard connectives, five modal operators with the following intuitive interpretations:

$\bigcirc\phi$	is interpreted as 'in any next instant ϕ will hold', dual is $\Diamond$;
$\bigcirc^-\phi$	'in the previous instant, ϕ did hold', dual is $\Diamond^-$;
$B\phi$	'the agent believes that ϕ';
$I\phi$	'the agent is informed that ϕ';
$A\phi$	'ϕ is true at every state', dual is E.

The formal semantics has a temporal part and an informational part:

The temporal part: a ***temporal frame*** is a tuple $\langle T, \rho \rangle$, where T is a non-empty set of 'instants', and $\rho : T \to T \cup \{\text{NULL}\}$ gives the temporal predecessor (or NULL when there is none), with the following condition imposed to eliminate cycles: for any $n \geq 1$, if $t = \rho^n(t')$ then $t' \neq t$.

The informational part: a ***temporal doxastic frame*** is a tuple

$$\langle T, \sigma, \Omega, \{\mathcal{B}_t, \mathcal{I}_t\}_{t \in T}\rangle,$$

where $\langle T, \sigma\rangle$ is a temporal frame and the $\mathcal{B}_t$ and $\mathcal{I}_t$'s are functions from Ω into 2^{Ω}. Ω is a non-empty set of *world states*, and $\mathcal{B}$ and $\mathcal{I}$ represent respectively the belief set and information received at a given state and instant. To get a temporal doxastic ***model*** (henceforth sometimes just 'model'), we add a valuation $V : \Phi \rightarrow 2^{\Omega}$, which says in which states the various proposition letters are true.

Given some model, a formula is evaluated with respect to a state and an instant – this is a 'two-dimensional' modal semantics. Thus we will define a satisfaction relation $\vDash \subseteq \Omega \times T \times \mathcal{L}_B$. We will write $[\![\phi]\!]_t$ to mean the set of states at which ϕ holds in the moment t, i.e. $\{\omega \in \Omega \mid \omega, t \vDash \phi\}$, and $[\![\phi]\!]^{\omega}$ to mean $\{t \in T \mid \omega, t \vDash \phi\}$. The interpretation of the modal operators in a model is then as follows:

$$\begin{array}{lll}
\omega, t \vDash \bigcirc\phi & \text{iff} & \rho^{-1}(t) \subseteq [\![\phi]\!]^{\omega} \\
\omega, t \vDash \bigcirc^{-}\phi & \text{iff} & \rho(t) = \text{NULL or } \omega, \rho(t) \vDash \phi \\
\omega, t \vDash B\phi & \text{iff} & \mathcal{B}_t(\omega) \subseteq [\![\phi]\!]_t \\
\omega, t \vDash I\phi & \text{iff} & \mathcal{I}_t(\omega) = [\![\phi]\!]_t \\
\omega, t \vDash A\phi & \text{iff} & \Omega = [\![\phi]\!]_t.
\end{array}$$

Let $\mathbb{L}_1$ be the logic (over $\mathcal{L}$) of temporal doxastic models.

THEOREM 2. $\mathbb{L}_1$ *is completely axiomatised by extending the axiomatisation for* varying-domain *temporal models of [Bon06] (where that logic is called* $\mathbb{L}_0$*') with the following six axiom schemata. (Note that* $P1$*–*$P4$ *are only for propositional variables, and that there is no replacement rule.)*

$$\begin{array}{llll}
\Diamond p \supset p & P1 & p \supset \bigcirc p & P2 \\
\Diamond^{-} p \supset p & P3 & p \supset \bigcirc^{-} p & P4 \\
\Diamond A\phi \supset A\Diamond\phi & D1 & \Diamond^{-} A\phi \supset A\Diamond^{-}\phi & D2
\end{array}$$

Proof. Along the lines of [Bon06], we build a *chronicle* [Bur84]. To remove defects we must add *families* of points defined with suitable care. ■

Bonanno defines a sub-class of models which he shows respect the AGM postulates *for Boolean* ϕ *and* ψ. The sense in which they respect the AGM postulates is spelled out in proposition 12 of [Bon07a], see also proposition 6 below. The restriction to pure Boolean ϕ and ψ is a non-trivial restriction, and parallels the restriction that we just discussed in section 1.1 with respect to the use of the AGM postulates in the artificial intelligence literature.

We will discuss the restriction further in section 3 below. Let Φ^B denote the pure Boolean formulae based on the set of proposition letters Φ. The members of Φ^B do not change their truth-conditions, i.e. the set of world states at which they are true:

REMARK 3 ([Bon07a], proposition 5). Let $\phi \in \Phi^B$. Fix an arbitrary model. Then, for every $t, t' \in T$, $[\![\phi]\!]_t = [\![\phi]\!]_{t'}$.

An ***AGM frame*** is a temporal doxastic frame satisfying (B1)–(B4).

$$\mathcal{B}_t(\omega) \subseteq \mathcal{I}_t(\omega) \tag{B1}$$

$$\text{If } \mathcal{B}_{\rho(t_1)}(\omega) \cap \mathcal{I}_{t_1}(\omega) \neq \emptyset \text{ then } \mathcal{B}_{t_1}(\omega) = \mathcal{B}_{\rho(t_1)}(\omega) \cap \mathcal{I}_{t_1}(\omega) \tag{B2}$$

$$\text{If } \mathcal{I}_t(\omega) \neq \emptyset \text{ then } \mathcal{B}_t(\omega) \neq \emptyset \tag{B3}$$

$$\text{If } \rho(t_1) = \rho(t_2),\ \mathcal{I}_{t_2}(\omega) \subseteq \mathcal{I}_{t_1}(\omega) \text{ and } \mathcal{I}_{t_2}(\omega) \cap \mathcal{B}_{t_1}(\omega) \neq \emptyset \text{ then } \mathcal{B}_{t_2}(\omega) = \mathcal{I}_{t_2}(\omega) \cap \mathcal{B}_{t_1}(\omega) \tag{B4}$$

In [Bon07b], Bonanno gives the formulae which characterise the defining properties of AGM frames (i.e. that together are valid precisely on AGM frames). The following section extends those results to postulates concerning iterated revision. We give further subclasses of frames which we show correspond to postulates for iterated revision in the same sense in which AGM frames correspond to the AGM postulates. We will also give modal axioms which characterise these classes.

2 Foreground: Iterated Revision

We will present a series of postulates for iterated revision. It is not our goal here to assess these postulates, but rather to give formal results to see how they fit into Bonanno's [Bon07a] temporal modal logic. We give semantic properties of frames which we show capture those postulates, and a syntactical characterisation of each of those properties in the form of axioms of the modal language $\mathcal{L}_B$.

2.1 Recalcitrance

Nayak et al [NPP03] introduce the following postulate of 'recalcitrance':

$$\text{If } \phi \nvdash \neg\psi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \vdash \phi \tag{$\dotplus$!}$$

That is, if you receive two pieces of consistent information one after the other, receipt of the second piece of information should not eliminate your belief in the first. We refer to [NPP03] for discussion of the merits of this constraint. On the semantic side, we define recalcitrant AGM frames:

DEFINITION 4. A ***recalcitrant*** AGM frame is an AGM frame in which

$$\text{If } \mathcal{I}_{\rho(t)}(\omega) \cap \mathcal{I}_t(\omega) \neq \emptyset, \text{ then } \mathcal{B}_t(\omega) \subseteq \mathcal{I}_{\rho(t)}(\omega) \tag{R}$$

Recalcitrant frames capture the postulate ($\dot{+}$!), a statement made formally by proposition 6. Furthermore, we give the syntactic correspondent of recalcitrance. Consider the following formula scheme, where $\phi, \psi \in \Phi^B$:

$$A\neg(\phi \wedge \psi) \vee (I\phi \supset \bigcirc(I\psi \supset B\phi)) \qquad \text{(REC)}$$

We say that a set of formula schemata S ***characterises a property*** P of AGM frames just when any AGM frame $\mathcal{F}$ has P iff every instance of S is valid on it.

PROPOSITION 5. REC *characterises recalcitrant AGM frames.*

Proposition 12 in [Bon07a] describes a connection between the class of AGM frames and the AGM postulates. In the same spirit, the following result connects the class of recalcitrant frames with ($\dot{+}$!).

PROPOSITION 6. REC *provides an axiomatic characterisation of recalcitrant revision, in the sense that (A) and (B) both hold*

(A) Let Γ be a belief state, and $\phi, \psi \in \Phi^B$ If Γ satisfies ($\dot{+}$!) with respect to ϕ and ψ, then there is a recalcitrant AGM model

$$\langle T, \sigma, \Omega, \{\mathcal{B}_t, \mathcal{I}_t\}_{t\in T}\rangle$$

with some $t_0, t_1, t_2 \in T, \omega \in \Omega$ such that

1. *$t_0 = \rho(t_1)$;*
2. *$\overline{\Gamma} = BTh(\omega, t_0)$;*
3. *$\omega, t_1 \vDash I\phi$;*
4. *$\overline{\Gamma \dot{+} \phi} = BTh(\omega, t_1)$;*
5. *If ϕ is consistent, then $\omega', t' \vDash \phi$ for some $\omega' \in \Omega, t' \in T$;*
6. *$t_1 = \rho(t_2)$;*
7. *$\omega, t_2 \vDash I\psi$;*
8. *$\overline{(\Gamma \dot{+} \phi) \dot{+} \psi)} = BTh(\omega, t_2)$;*
9. *If ψ is consistent, then $\omega'', t'' \vDash \phi$ for some $\omega'' \in \Omega, t'' \in T$;*

(B) Fix a recalcitrant AGM model such that (1) for some $t_0, t_1, t_2 \in T, \omega \in \Omega$ and $\phi, \psi \in \Phi^B$, $\rho(t_1) = t_0$, $\rho(t_2) = t_1$, $\omega, t_1 \vDash I\phi$, $\omega, t_2 \vDash I\psi$, (2) if ϕ is not a contradiction then $\omega', t' \vDash \phi$ for some $\omega' \in \Omega, t' \in T$, and (3) if ψ is not a contradiction then $\omega'', t'' \vDash \psi$ for some $\omega'' \in \Omega, t'' \in T$. Take any belief state Γ and function $\dot{+} : \underline{\Gamma \times \Phi^B \to} \Gamma$, such that $\overline{\Gamma} = BTh(\omega, t_0)$; $\overline{\Gamma \dot{+} \phi} = BTh(\omega, t_1)$ and $\overline{(\Gamma \dot{+} \phi) \dot{+} \psi}$. Then $\Gamma \dot{+} \phi$ and $(\Gamma \dot{+} \phi) \dot{+} \psi$ respect ($\dot{+}$!).

Furthermore, for every $\phi, \psi \in \Phi^B$*, there exists a recalcitrant AGM model such that, for some* $\omega \in \Omega$*,* $t_1, t_2 \in \Omega$*, (1)* $\omega, t_1 \vDash I\phi$ *and* $\omega, t_2 \vDash I\psi$*, (2) if* ϕ *is not a contradiction then* $\omega', t' \vDash \phi$ *for some* $\omega' \in \Omega$ *and* $t' \in T$ *and (3) if* ψ *is not a contradiction then* $\omega'', t'' \vDash \psi$*, for some* $\omega'' \in \Omega$ *and* $t'' \in T$*.*

We will give analogous results with respect to other constraints on iterated belief revision. When we say below that we have a ***semantic correspondent*** for an iterated belief revision postulate, we mean that we have a result analogous to proposition 6 with respect to that postulate.

2.2 Darwiche and Pearl

Darwiche and Pearl [DP97] propose four postulates for iterated belief revision, for which we provide here semantic and syntactic correspondents. We refer to [DP97] for motivation and explanation. The postulates are as follows:

$$\text{If } \psi \vdash \phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus \psi \qquad (\dotplus\text{DP1})$$

$$\text{If } \psi \vdash \neg\phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus \psi \qquad (\dotplus\text{DP2})$$

$$\text{If } \Gamma \dotplus \psi \vdash \phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \vdash \phi \qquad (\dotplus\text{DP3})$$

$$\text{If } \Gamma \dotplus \psi \nvdash \neg\phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \nvdash \neg\phi \qquad (\dotplus\text{DP4})$$

Each one of these postulates can be translated into semantic correspondents, i.e. properties of an AGM frame. The next proposition shows this.

PROPOSITION 7. *The following are semantic correspondents for (*$\dotplus$*DP1 –* $\dotplus$*DP4). That is, (DP1) is a correspondent for (*$\dotplus$*DP1), and so forth. (*) abbreviates* $\rho(t_0) = \rho(\rho(t_1))$ *and* $\mathcal{I}_{t_0}(\omega) = \mathcal{I}_{t_1}(\omega)$*.*

$$\textit{If (*) and } \mathcal{I}_{t_0}(\omega) \subseteq \mathcal{I}_{\rho(t_1)}(\omega)\textit{, then } \mathcal{B}_{t_0}(\omega) = \mathcal{B}_{t_1}(\omega) \qquad \text{(DP1)}$$

$$\textit{If (*) and } \mathcal{I}_{t_0}(\omega) \cap \mathcal{I}_{\rho(t_1)}(\omega) = \emptyset\textit{, then } \mathcal{B}_{t_0}(\omega) = \mathcal{B}_{t_1}(\omega) \qquad \text{(DP2)}$$

$$\textit{If (*) and } \mathcal{B}_{t_0}(\omega) \subseteq \mathcal{I}_{\rho(t_1)}(\omega)\textit{, then } \mathcal{B}_{t_1}(\omega) \subseteq \mathcal{I}_{\rho(t_1)}(\omega) \qquad \text{(DP3)}$$

$$\textit{If (*) and } \mathcal{B}_{t_0}(\omega) \cap \mathcal{I}_{\rho(t_1)}(\omega) \neq \emptyset\textit{, then } \mathcal{B}_{t_1}(\omega) \cap \mathcal{I}_{\rho(t_1)}(\omega) \neq \emptyset \qquad \text{(DP4)}$$

Furthermore, each of these semantic properties is characterised syntactically by an $\mathcal{L}_B$ formula:

PROPOSITION 8. *The following formula schemata (*$\phi, \psi, \chi \in \Phi^B$*) characterise (DP1–DP4)*

$$(A(\psi \supset \phi) \wedge \Diamond(I\psi \wedge B\chi)) \supset \bigcirc(I\phi \supset \bigcirc(I\psi \supset B\chi)) \qquad \text{(D1A)}$$

$$(A(\psi \supset \phi) \wedge \Diamond(I\phi \wedge \Diamond(I\psi \wedge B\chi))) \supset \bigcirc(I\psi \supset B\chi) \qquad \text{(D1B)}$$

$$(A(\psi \supset \neg\phi) \wedge \Diamond(I\psi \wedge B\chi)) \supset \bigcirc(I\phi \supset \bigcirc(I\psi \supset B\chi)) \qquad \text{(D2A)}$$

$$(A(\psi \supset \neg\phi) \wedge \Diamond(I\phi \wedge \Diamond(I\psi \wedge B\chi))) \supset \bigcirc(I\psi \supset B\chi) \qquad \text{(D2B)}$$

$$\Diamond(I\psi \wedge B\phi) \supset \bigcirc(I\phi \supset \bigcirc(I\psi \supset B\phi)) \qquad \text{(D3)}$$

$$\Diamond(I\psi \wedge \neg B\neg\phi) \supset \bigcirc((I\phi \wedge \Diamond I\psi) \supset \Diamond(I\psi \wedge \neg B\neg\phi)) \qquad \text{(D4)}$$

2.3 Refinement

A very natural revision procedure, at least when one approaches belief revision from a semantic perspective using plausibility pre-orders à la Grove [Gro88], is that of lexicographic re-ordering. This policy is discussed under that name ('lexicographic') by Nayak et al [NPP03], and also by van Benthem [Ben06]. Baltag and Smets [BS07] take the lexicographic process to be a primitive in their semantics. In Rott's taxonomy [Rot06] it is called the 'moderate' policy. We here call it 'refinement', a name taken from Shoham and Maynard-Reid II in a different context [MS01]. Refinement is a strengthening of the Darwiche and Pearl postulates. AGM-style postulates are given in [Zve07] for refinement as follows:

$$\text{If } \psi \vdash \neg\phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus \phi \qquad (\dotplus\text{R1})$$

$$\text{If } \psi \nvdash \neg\phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus (\phi \wedge \psi) \qquad (\dotplus\text{R2})$$

Notice that ($\dotplus$R1) is just ($\dotplus$DP2), so we only need results for ($\dotplus$R2). Although it is not our goal to motivate any particular policy here, we can note that this policy is extremely natural mathematically, and furthermore that ($\dotplus$R2) has some intuitive appeal: If two incoming pieces of information are consistent, then accepting one and then the other is the same as accepting both at the same time. *Passons.* The formal results:

PROPOSITION 9. *The semantic correspondent for ($\dotplus$R2) is:*

If $\rho(t_0) = \rho(\rho(t_1))$ and $\mathcal{I}_{t_0}(\omega) = \mathcal{I}_{\rho(t_1)}(\omega) \cap I_{t_1}(\omega) \neq \emptyset$, then $\mathcal{B}_{t_1}(\omega) = \mathcal{B}_{t_0}(\omega)$. (R2)

PROPOSITION 10. *The following formula schemata (with $\phi, \psi, \chi \in \Phi^B$) characterise the property R2:*

$$A\neg(\phi \wedge \psi) \vee (\Diamond(I(\phi \wedge \psi) \wedge B\chi) \supset \bigcirc(I\phi \supset \bigcirc(I\psi \supset B\chi))) \qquad \text{(R2A)}$$

$$A\neg(\phi \wedge \psi) \vee (\Diamond(I\phi \wedge \Diamond(I\psi \wedge B\chi)) \supset \bigcirc(I(\phi \wedge \psi) \supset B\chi)) \qquad \text{(R2B)}$$

2.4 Scepticism

The other policy which van Benthem [Ben06] considers and axiomatises he calls 'elite' or 'Machiavellian' revolution. (Darwiche and Pearle also present it [DP97], calling it 'absolute minimisation'. We take the name 'scepticism' from Segerberg [Seg]). In this policy, only the 'elite' states are upgraded, so that should one receive two pieces, ϕ then ψ, of information, such that ψ contradicts the belief *set* established after revision by ϕ, then the agent will, 'skeptically', act as if ϕ had never been received.

Darwiche and Pearl [DP97] give a postulate for this revision policy:

$$\text{If } \Gamma \dotplus \phi \vdash \neg\psi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus \psi \qquad (\dotplus\text{S})$$

We have semantic and syntactic correspondents for this postulate in $\mathcal{L}_B$:

PROPOSITION 11. *($\dotplus$S) corresponds semantically to:*

If $\rho(t_0) = \rho(\rho(t_1))$, $\mathcal{I}_{t_1}(\omega) = \mathcal{I}_{t_0}(\omega)$, *and* $\mathcal{B}_{\rho(t_1)}(\omega) \cap \mathcal{I}_{t_1}(\omega) = \emptyset$ *then* $\mathcal{B}_{t_1}(\omega) = \mathcal{B}_{t_0}(\omega)$. (S)

PROPOSITION 12. *The following formula schemata (with $\phi, \psi, \chi \in \Phi^B$) characterise the property S:*

$$\Diamond(I\psi \wedge B\chi) \supset \bigcirc(I\phi \supset (B\neg\psi \supset \bigcirc(I\psi \supset B\chi))) \qquad (\textsc{sa})$$

$$\Diamond(I\psi \wedge B\neg\psi \wedge \Diamond(I\psi \wedge B\chi)) \supset \bigcirc(I\psi \supset B\chi) \qquad (\textsc{sb})$$

2.5 Cynicism

All of the previous policies are presented in Rott's [Rot06] extensive taxonomy. The policy that we present next is not. It is a strengthening of the skeptically policy, which we call the 'cynical' policy. Following van Benthem's metaphor, we might have called it the 'backstabbing Machiavellian' policy, because here with a revision by ϕ first of all only the *top* ϕ states are upgraded, and furthermore the other ϕ states are actually *downgraded*. An agent following the cynical policy does accept incoming information, but should she discover some new information which does not entail the old information and which combined with it leads her to have a false belief, she will suppose the old information to have been a lie, and will actually believe the opposite of it. If the new information does entail the old information, she simply revises by the new information. [Zve07] introduces the policy, for which we give the following AGM-style postulates:

$$\text{If } \Gamma \dotplus \phi \vdash \neg\psi \text{ and } \psi \nvdash \phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus (\neg\phi \wedge \psi) \qquad (\dotplus\text{C1})$$

$$\text{If } \psi \vdash \phi \text{ then } (\Gamma \dotplus \phi) \dotplus \psi \equiv \Gamma \dotplus \psi \qquad (\dotplus\text{C2})$$

Notice that ($\dotplus$C2) is just (DP1), so it suffices to give a semantic correspondent and a syntactic characterisation of ($\dotplus$C2):

PROPOSITION 13. *($\dotplus$C1) corresponds semantically to:*

If $\rho(t_0) = \rho(\rho(t_1))$, $\mathcal{I}_{t_0}(\omega) = \mathcal{I}_{t_1}(\omega) - \mathcal{I}_{\rho(t_1)}(\omega) \neq \emptyset = \mathcal{B}_{\rho(t_1)}(\omega) \cap \mathcal{I}_{t_1}(\omega)$, *then* $\mathcal{B}_{t_1}(\omega) = \mathcal{B}_{t_0}(\omega)$. (C)

PROPOSITION 14. *The following formula schemata (with $\phi, \psi, \chi \in \Phi^B$) characterise the property (C):*

$$A(\psi \supset \phi) \vee (\Diamond(I(\psi \wedge \neg\phi) \wedge B\chi) \supset \bigcirc(I\phi \wedge B\neg\psi) \supset \bigcirc(I\psi \supset B\chi)) \qquad \text{(CA)}$$

$$\Diamond(I\phi \wedge \Diamond(I\psi \wedge B\chi)) \supset \bigcirc(I(\psi \wedge \neg\phi) \supset B\chi) \qquad \text{(CB)}$$

3 Shadow: Beyond Boolean formulae

As we have seen, Bonanno [Bon07a] restricts the formulae by which it is possible to revise to pure Boolean. We have done the same in our characterisation results for iterated revision. This reflects the AGM postulates, which are concerned with factual change. However, looking a little beyond the AGM paradigm, it is a significant restriction, and a particular impediment to extending the logic to reasoning about the multi-agent case (with the concomitant applications to reasoning about extensive games which the branching time semantics suggests).

If the restriction to pure Boolean formulae is dropped naïvely without anything being changed, then in the one-agent case, Moore problems arise otherwise due to the success postulate ($\dotplus$1): if we allow revision by a Moore sentence $\phi := p \wedge \neg Bp$, then we have the unappealing consequence that after such a revision, the agent believes $(p \wedge \neg Bp)$: If we want agents to be positively introspective, then we will have a contradiction. (Similar examples can be constructed in a multi-agent version of this logic, which would be of interest in analysing interactions.)

Thus having answered one research question of Bonanno's, we would like to raise a new question: Is it possible in this branching time semantics to retain the spirit of the AGM postulates while throwing away the letter, obtaining a coherent treatment of the Moore sentences and a coherent extension to multi-agent belief revision?

The Moore problems and their multi-agent counterparts are handled well by taking a dynamic epistemic logic [BMS99] approach to belief revision, as proposed and developed for belief revision by van Benthem [Ben06] and Baltag and Smets [BS07]. Bonanno has elsewhere [Bon05] suggested that it is a flaw in a modal logic that it should contain a modal operator for each formula. In the logic proposed by van Benthem there is a *binary* modality, for the static semantics which involves pre-orders, and a monadic modality for dynamic revision.[6] It should be clear that the formulae $\langle \dotplus\phi \rangle \psi$ and $\Diamond(I\phi \wedge \psi)$ have the same intended meaning, viz. 'it's possible to revise by ϕ and have ψ hold'. Therefore it would be of interest to see spelled out more

[6] van Benthem [Ben06] gives 'reduction axioms' for the two policies refinement and scepticism; note that for anything short of a policy no reduction axiom would possible.

clearly the precise objection against logics which are not parsimonious with their modalities. In any case, in what follows we will suggest a way to exploit the temporal structure of the models in order to remove the restriction to pure Boolean formulae.

Yap [Yap06] introduces a Past modality to dynamic epistemic logic, a generalisation of public announcement logic (PAL, [Pla89]). She makes the remark that after a public announcement of ϕ, 'what actually becomes common knowledge is not ϕ, but that ϕ was true just before the announcement' (op. cit. p. 11). [Zve07] makes similar observations with respect to belief revision in the context of Segerberg's dynamic doxastic logic [Seg95]. The approach that we now propose is to exploit the temporal structure of the models under consideration in order to take this analysis further.

We will modify the way we interpret the models, representing 'after the agent revises by ϕ, ψ holds' by $I\phi \supset \bigcirc\psi$ (up until now this would have been $I\phi \supset \psi$). That is, we say that when a piece of information is received, action (revision for example) is taken on it in the *next* step. This is important because it allows a simple representation of incoming information that describes the state of the world *before the information arrived.* For reasons of clarity, we will focus on the simplest axiom (A) from $\mathbb{L}_{AGM}$, which corresponds to (B1), the semantic correspondent of the success postulate ($\dot{+}1$). Nonetheless the insights remain relevant for the other axioms.

$$I\phi \supset B\phi \tag{A}$$

The new way of interpreting models entails replacing (A) with (A$^+$):

$$I\phi \supset \bigcirc B \bigcirc^- \phi \tag{A$^+$}$$

(A$^+$) can be read '*after* receiving some piece of information, the agent believes that ϕ *was* the case.' We propose that this way of presenting the impact of information is natural, and is certainly suited to the case of an agent receiving information about the world. We give formal force to this reading of (A$^+$) with the following proposition:

PROPOSITION 15. *(A$^+$) characterises this property:* $\mathcal{B}_t(\omega) \subseteq \mathcal{I}_{\rho(t)}(\omega)$

In the rest of this section we will restrict ourselves to sketching a way to treat in these branching time models of information change the simpler case of belief '*expansion*', i.e. where incoming information is always consistent with existing information. This is precisely the case that is well-treated by PAL. We will show that our suggestion for modelling belief dynamics in these temporal models bears many similarities to the PAL model.

To treat the PAL case, we extend the base logic $\mathbb{L}_1$ for temporal doxastic frames with the following axiom schemata:

$$I\phi \supset AI\phi \tag{UA}$$

$$I\phi \supset (\Diamond B \Diamond^{-} \psi \supset B(\phi \supset \psi)) \tag{NM}$$

$$I\phi \supset (B(\phi \supset \psi) \supset \bigcirc B \bigcirc^{-} \psi) \tag{PR}$$

(UA) stands for 'uniform announcement', saying that the same announcement is made at every point in the model. We include it for the sake of making the comparison with PAL more conspicuous. (NM) stands for 'no miracles', (PR) for 'perfect recall'.[7] We will define a class of temporal doxastic frames to which these axioms correspond. The axioms (NM) and (PR) are similar to the key 'reduction axiom' for the belief modality for PAL. In PAL, when an announcement of ϕ is made, the model is reduced by eliminating the states where ϕ is false. An alternative approach is to say rather that only the epistemic *relation* is reduced, so that the states survive (cf. [OR94], p. 72). It is this approach that we can mimic in this branching next-time semantics, because the domain Ω is constant across instants.

We will define a class of frames in which information flows in the appropriate public announcement manner. Specifically, we want that when an agent receives the information that ϕ, in the next state she will eliminate from consideration all the world states at which ϕ did hold. This is precisely what the following definition gives us:

DEFINITION 16. ***Public announcement*** temporal doxastic frames are those in which the following properties hold:

$$\mathcal{I}_t(\omega) = \mathcal{I}_t(\omega') \quad \text{(so we can just write } \mathcal{I}_t\text{)} \tag{PA1}$$

$$\text{if } \rho(t) \neq \text{NULL then } \mathcal{B}_t(\omega) = \mathcal{I}_{\rho(t)}(\omega) \cap \mathcal{B}_{\rho(t)}(\omega) \tag{PA2}$$

PROPOSITION 17. $(UA) \wedge (NM) \wedge (PR)$ *characterises public announcement temporal doxastic frames.*

The information flow in these frame mirrors that of PAL, as we will show with an example, in which we use multi-agent frames, i.e. with not just one belief accessibility relation $\mathcal{B}$ but a family of them $\mathcal{B}_a$, and a language with a modality for each of them, with instances of (NM) and (PR) for each. A puzzle familiar from the literature, of the 'muddy children', or the 'hats' [Ben07a, GG97, OR94] can be formulated and resolved in the resulting logic. Figure 1 shows a public announcement temporal doxastic model for the simple two-child case. (We use solid dots to represent states which are self-accessible for both relations, and empty dots to represent those which are self-accessible for neither.) To those familiar with one of the

[7] van Benthem and Pacuit [BP06] use this vocabulary in a slightly different context to describe properties of agents in temporal models.

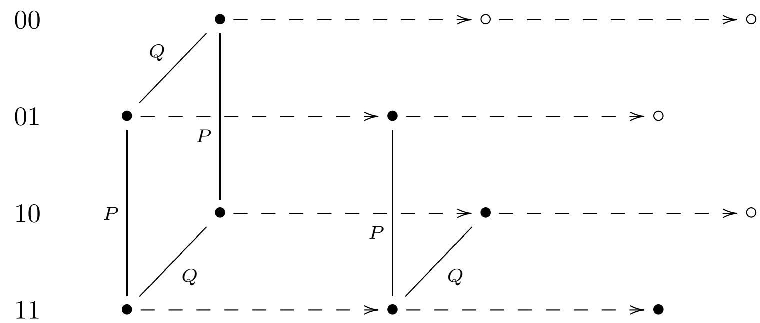

Figure 1. A model of a simple muddy children puzzle

many appearances of this puzzle in the dynamic epistemic logic literature [Ben03, Ben07b], our representation of it will be recognisable. Two children Patricia and Quentin can only see the forehead of the other, and are wondering who has mud on their forehead. B_p is Patricia's, and B_q Quentin's, belief modality; p and q are propositional variables interpreted to mean that Patricia and Quentin, respectively, are muddy. Four world states are relevant for them: Neither is muddy 00, only Patricia is muddy 10, only Quentin is muddy 01, or both are muddy 11. That epistemic situation is represented by the accessibility relations in the model at the first instant t_0, when they are told that at least one of them is muddy ($\mathcal{I}_{t_0} = [\![p \vee q]\!]_{t_0}$). Then at a next instant t_1, they learn that neither of them knows whether he or she is muddy ($\mathcal{I}_{t_1} = [\![\neg(B_p p \vee B_p \neg p) \wedge \neg(B_q q \vee B_q \neg q)]\!]_{t_1}$). Because this is a public announcement temporal doxastic model, the belief relations at t_1 and any t_2 such that $\rho(t_2) = t_1$ are determined. Similarly in the PAL case, given the initial static model and an announcement, the new static model is determined. The axioms that characterise this class of models enable us to derive for example that if $B_p(\neg p \supset B_q \neg p)$, as is the case at t_0, then after these announcements, at t_2, $B_p p$, i.e. Patricia knows that she is muddy.

Public announcement temporal doxastic models describe what van Benthem et al [BGP07] call 'protocol sets' as opposed to PAL's 'full protocols': rather than allowing *every* truthful event to occur, they specify which informative events can occur.

So much for the connection with public announcement logic. What about belief revision? We have shown how to use the temporal modalities to get a dynamic version (A^+) of (A). To give dynamic versions of the other static formulae and properties given in [Bon07b] connecting branching time models to the AGM postulates is not an entirely trivial exercise.

We leave a development of the direction indicated here for future work. We also postpone the question of a complete axiomatisation of the logic of public announcement temporal doxastic frames, including any possibilities for axiomatising common knowledge (cf. [BP06, Yap06, BEK05]).

Conclusion

We have answered a research question posed by Bonanno [Bon07a], by giving axiomatic characterisations, for branching time models, of several constraints on iterated revision. We gave a complete axiomatisation of the logic of those models.

We have also introduced a new question concerning the pure Boolean nature of incoming information, seeing what happens when that information can be about beliefs or time, and we suggested an approach that would answer this question. Specifically, we endorsed a recognition of the 'dynamic' nature of epistemic events [Ben06], and proposed giving it substance in these branching next-time models by exploiting the temporal structure. We have shown that in the simple case of belief *expansion*, the approach we propose works along the same lines as public announcement logic. We claim that this supports our modification of Bonanno's framework to handle dynamic revision as well as we have shown it to handle iterated revision.

Acknowledgements

I thank Giacomo Bonanno for useful feedback on parts of an earlier version of this paper, and Johan van Benthem and Krister Segerberg for introducing me to formal theories of belief revision. Andreas Witzel pointed out some errors in an early draft; those remaining are my responsibility.

BIBLIOGRAPHY

[AGM85] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *JSL*, 50(2):510–530, 1985.

[BEK05] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Common knowledge in update logics. In *TARK '05: Proceedings of the 10th conference on Theoretical aspects of rationality and knowledge*, pages 253–261, Singapore, Singapore, 2005. National University of Singapore.

[Ben03] Johan van Benthem. One is a lonely number. *ILLC Prepublication*, PP-2003(07), 2003.

[Ben06] Johan van Benthem. Dynamic logic for belief revision. *ILLC Prepublication*, PP-2006(11), 2006.

[Ben07a] Johan van Benthem. Actions that make us know. *ILLC Prepublication*, PP-2007(10), 2007.

[Ben07b] Johan van Benthem. Rational dynamics and epistemic logic in games. to appear in International Journal of Game Theory, 2007.

[BGP07] Johan van Benthem, Jelle D. Gerbrandy, and Eric Pacuit. Merging frameworks for interaction: DEL and ETL. accepted for TARK XI, 2007.

[BMS99] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge and private suspicions. Technical Report SEN-R9922, Centrum voor Wiskunde en Informatica, 1999.

[Bon05] Giacomo Bonanno. A simple modal logic for belief revision. *Synthese*, 147(2):193–228, 2005.

[Bon06] Giacomo Bonanno. A sound and complete temporal logic for belief revision. Manuscript, Dec. 2006.

[Bon07a] Giacomo Bonanno. Axiomatic characterization of the AGM theory of belief revision in a temporal logic. *Artificial Intelligence*, 171(2–3):144–160, 2007.

[Bon07b] Giacomo Bonanno. Belief revision in a temporal framework. Manuscript, Feb. 2007.

[Bou96] Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *JPL*, 25:263–305, 1996.

[BP06] Johan van Benthem and Eric Pacuit. The tree of knowledge in action: Towards a common perspective. In Guido Governatori, Ian M. Hodkinson, and Yde Venema, editors, *Advances in Modal Logic*, pages 87–106. College Publications, 2006.

[BS06] Alexandru Baltag and Mehrnoosh Sadrzadeh. The algebra of multi-agent dynamic belief revision. *Electronic Notes in Theorerical Computer Science*, 157(4):37–56, 2006.

[BS07] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. Manuscript, 2007.

[Bur84] John P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II, pages 89–133. Reidel, 1984.

[DP97] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1–2):1–29, 1997.

[GG97] Jelle D. Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.

[Gro88] Adam Grove. Two modellings for theory change. *JPL*, 17(2):157–170, 1988.

[LHM95] Bernd van Linder, Wiebe van der Hoek, and John-Jules Ch. Meyer. Actions that make you change your mind. In *KI '95: Proceedings of the 19th Annual German Conference on Artificial Intelligence*, pages 185–196, London, UK, 1995. Springer-Verlag.

[Lin95] Bernd van Linder. A dynamic logic of iterated belief change. Technical Report UU-CS-1995-40), Utrecht University, 1995.

[MS01] Pedrito Maynard-Reid II and Yoav Shoham. Belief fusion: Aggregating pedigreed belief states. *Journal of Logic, Language, and Information*, 10(2):183–209, 2001.

[NPP03] Abhaya C. Nayak, Maurice Pagnucco, and Pavlos Peppas. Dynamic belief revision operators. *Artificial Intelligence*, 146(2):193–228, 2003.

[OR94] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. MIT, 1994.

[Pla89] Jan A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.

[Rot06] Hans Rott. Shifting priorities: Simple representations for twenty-seven iterated theory change operators. In Henrik Lagerlund, Sten Lindström, and Rysiek Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, volume 53 of *Uppsala Philosophical Studies*, pages 359–384. Uppsala, 2006.

[Seg] Krister Segerberg. Iterated belief revision in dynamic doxastic logic. Manuscript.

[Seg95] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the Interest Group in Pure and Applied Logics*, pages 535–553, 1995.

[Seg06] Krister Segerberg. Moore problems in full dynamic doxastic logic. In Jacek Malinowski and Andrzej Pietruszczak, editors, *Essays in Logic and Ontology*, volume 91 of *Poznan Studies in the Philosophy of the Sciences and the Humanities*, pages 95–110. Rodopi, November 2006.

[Yap06] Audrey Yap. Product update and looking backward. *ILLC Prepublication*, PP-2006(39), 2006.

[Zve07] Jonathan A. Zvesper. Belief revision and epistemic acts. Master's thesis, Institute for Logic, Language and Computation, Amsterdam, 2007. ILLC Master of Logic thesis MoL-2007-01.