A model for updates in a multi-agent setting

John Cantwell

May 26, 2005

1 Introduction

This paper deals with the problem of how to represent changes in belief when several agents are witnesses to some event in the world and they are witnesses to the event in such a manner that it is common knowledge that they have all witnessed it. The problem is closely analogous to the problem of how to represent changes in belief when common announcements are made, announcements where it becomes common knowledge among the agents that the announcement is true and that an announcement has been made.

Katsuno and Mendelzon's (Katsuno and Mendelzon, 1992) (henceforth KM) is the first formal representation of the operation of *updating* a belief state that explicitly makes a distinction between an update and a *revision*. In a revision with a proposition A one goes from a state of disbelief or non-belief in A, to a belief in A; it is the operation of finding out that something is true that one previously did not believe was true. In an update with A one finds out that something has changed, that A has *become* true.

KM uses a semantics that is formally equivalent to David Lewis' semantics for subjunctive (e.g. counter-factual) conditionals. Thus it is possible to re-interpret their idea of an update in terms of (subjunctive/counter-factual) *suppositional* reasoning: the same formal framework can be interpreted both as answering the question "A has become true, so what is now the case?" and the question "suppose that A *were/had been/would become* true, what would then *be/have been* the case?".[1] There is a close analogy here in the relationship between belief revision and (indicative) suppositional reasoning. As Levi (Levi, 1996) has pointed out, the AGM framework for belief revision can be interpreted as a formal representation intended to answer questions of the form "suppose that A is true, what is then the case?".

In the framework studied by KM a distinction is made between the belief-independent features of the world, and the belief state of the agent. Only one agent is considered. KM thinks of an update as applying to changes in the belief-independent features of the world, such as: finding out that a door has been opened, finding out that someone has moved the green box on the blue

[1] Note also that in Causal Decision Theory (Joyce, 1999), the relationship between subjunctive conditionals and the consequences of action is central.

box (where it previously was on the yellow box), finding out that someone has made a move in a game, and so on. Of course, after an update has been made an additional change in the world has occurred: the belief state of the agent has changed. In KM's framework the agent has no beliefs about herself and as there are no other agents involved in the framework, a change in the belief-dependent features of the world need not be representable by an update.

In a typical game, however, there are at least two agents involved and they not only both observe that the opponent has made a move, they often make this observation in a way that makes it common knowledge that the move has been made. Thus, not only must they update their belief states about the belief-independent features of the world, they must update their belief states knowing that the opponent has updated her belief state as well, both about the fact that the move has been made but also about the fact that the opponent's opponent has updated her belief state. Thus a move in a game is a change that brings about a change in the players' belief states, and these latter changes must also be representable by the update operation.

The model for updates in multi-agent settings that will be presented here builds on a formal model presented in (Cantwell, 2005) (for alternative frameworks see (Baltag et al., 1999; Baltag and Moss, 2004) in particular). The model developed there was as a framework to handle common announcements. Briefly, a common announcement is a situation in which several agents *revise* (not *update*) with a proposition P in such a way that it becomes common knowledge that they all revise with P. There are several alternative frameworks to handle common announcements (e.g. (Plaza, 1989; Fagin et al., 1995; Gerbrandy and Groeneveld, 1997; Gerbrandy, 1999; Baltag et al., 1999; Moss, 1999; Kooi, 2003; Baltag and Moss, 2004)) but the model presented in (Cantwell, 2005) has three advantages.

First, it provides a *modular* representation of possible worlds: a world is represented as an $n+1$-tuple $\langle u, b_1, b_2, \ldots, b_n\rangle$ where u determines the belief independent features of the world, and b_1 through b_n represent the belief states of the agents 1 to n. The advantage of the modular approach is that a local change in, say, the belief state of agent 1, can be represented as a transition from $\langle u, b_1, b_2, \ldots, b_n\rangle$ to $\langle u, b'_1, b_2, \ldots, b_n\rangle$ where all that has changed is the belief state of agent 1.

The second advantage of the framework presented in (Cantwell, 2005) is that it does not, unlike the modular approach found in (Gerbrandy and Groeneveld, 1997), rely on non-wellfounded set theory; ordinary set theory will do. This is obtained by letting the belief states b_i be entities in their own right and not sets of possible worlds. Belief states are, however, *associated* with a set of possible worlds by the function $\mathcal{W}(b_i)$ (which returns the set of possible worlds consistent with the belief state b_i). In (Cantwell, 2005) it was shown that this framework, at the level of logic, is equivalent to the framework of (Gerbrandy and Groeneveld, 1997).

The third advantage, connected with the first two, is that the framework is amenable to the generalization of intuitions about the structure of belief changing operations from the non-introspective single-agent setting, as well as the formal structures that have been developed to model such operations. So, for instance, the AGM (Alchourrón et al., 1985) selection function designed to model non-introspective single-agent revision turns out to be immediately applicable to the case of common announcements.

This paper presents a formal framework for multi-agent updates. As noted, the semantics builds on the ideas in (Cantwell, 2005). The formal language will contain doxastic operators, dynamic operators (thus it is a species of DDL - Dynamic Doxastic Logic (Segerberg, 1996)) and counter-factual conditionals. An axiomatisation is given which is both sound and complete.

2 The n-agent epistemic framework

Denote the agents by a number $1, \ldots, n$. Let U be a set of belief-independent states of the world, and $\mathcal{B}_i$ be the set of possible belief-states for the agent i. The belief states are taken to be entities in their own right (they are not sets of possible worlds). A *possible world* is an ordered $n+1$-tuple $\langle u, b_1, \ldots, b_n \rangle$ where $u \in U$ and, for each i, $b_i \in \mathcal{B}_i$. Let W be the set of possible worlds. Let $\mathcal{W}$ be a function that returns, for any agent i and belief state $b \in \mathcal{B}_i$, a set of possible worlds ($\mathcal{W}(b)$ can be seen as giving the *content* of the belief state b, the set of possible worlds that b has not yet excluded).

If w is a world $\langle u, b_1, \ldots, b_n \rangle$, let:

$$\begin{aligned} \mathrm{wst}(w) &= u \quad \text{(this gives the "world-state" of the world } w\text{)}, \\ \mathrm{bst}_i(w) &= b_i \quad \text{(this gives the belief state of agent } i \text{ in world } w\text{)}. \end{aligned}$$

I will assume *full introspection*:

$$\text{If } b \in \mathcal{B}_i \text{ and } w \in \mathcal{W}(b), \text{ then } \mathrm{bst}_i(w) = b.$$

Definition 1 *An* n-agent frame *is a tuple $\langle W_P, U, \mathcal{B}, \mathcal{W} \rangle$. Here W_P is a boolean set-algebra with W as its top element and $\emptyset$ as its bottom element; the elements of W_P will be referred to as propositions.*

In (Cantwell, 2005) it was shown that the n-agent frame is equivalent to standard Kripke semantics with n accessibility relations.

3 Similarity functions

For any proposition $P \in W_P$, let $\mathrm{sim}_w(P)$ denote a subset of P. Following David Lewis (Lewis, 1973), $\mathrm{sim}_w(P)$ can be thought of as returning those P-worlds that are most similar to w. A different interpretation, along the lines of (Katsuno and Mendelzon, 1992), would be to view $\mathrm{sim}_w(P)$ as giving the result of P becoming true (P 'happening') when one is in the world w. On this interpretation, the result of P becoming true can be non-deterministic: when $\mathrm{sim}_w(P)$ contains more than one world.

Only one restriction will here be imposed on sim:

(Weak Centering) $w \in \text{sim}_w(P)$, if $w \in P$.

To compress the notation somewhat, when b is a belief state, let $\text{sim}_b(P) = \bigcup\{\text{sim}_w(P)|w \in \mathcal{W}(b)\}$. $\text{sim}_b(P)$ can be seen as giving the *image* of the (worldly contents of the) belief state b under P: it is the set of possible ways of P becoming true that are consistent with the agent's beliefs about the initial state of the world.

4 Updates: The one-agent case

In this section updates will be studied for the one-agent case. This case is not entirely trivial as the agent has beliefs about her own beliefs, so when her first order beliefs change, her higher-order beliefs must follow suit.

An *update* with P is the result of finding out that P has become true. Updates are epistemic operations that do not change the belief-independent features of the world (so one can update with P even though P hasn't actually become true). Let $b \oplus P$ be the function that returns the belief state of the agent (initially in belief state b) upon having updated with P.

As the belief state of the agent is part of the world, the result of an update will be a change in the world. The change in the world will be represented by the function $\oplus(w, P)$ that returns the world that is the result of the agent in w having updated with P:

Definition 2 *For any world w and proposition P, let*

$$\oplus(w, P) = \langle wst(w), bst_1(w) \oplus P\rangle \text{ (recall that there is only one agent).}$$

Note that the function $\oplus(w, P)$ is fully determined by $b \oplus P$ and so is not an additional primitive notion. So: what is the new belief state that results from updating with P? That is, how is the function $b \oplus P$ defined?

Definition 3 $\oplus$ *is an* update function *relative an 1-agent frame $\mathcal{F}$ and similarity function sim if and only if for any $P \in W_P$ and any $b \in \mathcal{B}_1$:*

$$\mathcal{W}(b \oplus P) = \{\oplus(w, P)|w \in sim_b(P)\}$$

The definition takes the form of a fixed-point constraint, for spelling out the right hand side we get:

$$\mathcal{W}(b \oplus P) = \{\langle \text{wst}(w), \text{bst}(w) \oplus P\rangle|w \in \text{sim}_b(P)\}$$

To see what is going on here note that the 'first-order' or 'factual beliefs' (beliefs like "it is raining") of the agent after updating with P will be the result of imaging the agent's original belief state with P using the similarity function. That is, every world w in the original belief state has been mapped to $\text{sim}_w(P)$ and the new set of first-order beliefs will be determined by the set $\{\text{wst}(w)|w \in \text{sim}_b(P)\}$. But the worlds in $\text{sim}_w(P)$ need not be worlds in which the agent believes that P is true (for P might become true without the agent knowing it). So if the new belief state supported the worlds $\mathrm{sim}_b(P)$ the agent would have the first order belief that P, but not necessarily the higher order belief that the agent believes that P, which would violate the requirement of introspection.

To satisfy the requirement of introspection, the agent must along with the first order beliefs $\{\mathrm{wst}(w)|w \in \mathrm{sim}_b(P)\}$, have the second order beliefs $\{\mathrm{bst}(w) \oplus P|w \in \mathrm{sim}_b(P)\}$.

Say that, originally, the actual world was w. After P has become the case the actual world is now one of $\mathrm{sim}_w(P)$. So after updating with the fact that P has become true, one of the worlds $\{\oplus(w', P)|w' \in \mathrm{sim}_w(P)\}$ will be the actual world. If the agent knew, originally, that w was the actual world (so that $\mathcal{W}(\mathrm{bst}_1(w)) = \{w\}$), the new belief state $(b \oplus P)$ will, according to definition 5, support the set of worlds $\{\oplus(w, P)|w \in \mathrm{sim}_b(P)\}$. So if the agent knew, originally, what the actual world was like, her belief state after the update will be the set of worlds that could possibly result from P becoming true and the agent updating with this fact (i.e. the agent's beliefs will all be true, but if P is a non-deterministic change the agent will not know which of the possible ways that P has been realized).

5 Updates: The multi-agent case

To deal with multiple agents that update in such a way that it becomes common knowledge that the agents have updated, the definitions of the previous section need to be generalized.

For a belief state $b \in \mathcal{B}_i$, a proposition P and set of agents N (containing i), let $b \oplus_N P$ be the belief state of agent i that results from updating with the fact that P in such a way that the agent believes that it is common knowledge among the agents in N that they all have updated with P. Note that as the proposition P is a set of possible worlds (and not a set of belief-independent states) it is quite possible for the agent to update with the information that some other agent's belief state has changed: updates in this framework are not confined to changes in the belief-independent state of the world.

The function $\oplus(w, P, N)$ is the world that results when, in the world w, all the agents in N update with P.

Definition 4 *For any world w, proposition P and set of agents N, let $\oplus(w, P, N)$ denote the world w' where:*

1. *$wst(w') = wst(w)$,*
2. *$bst_i(w') = bst_i(b) \oplus_N P$ for each $i \in N$, and*
3. *$bst_i(w') = bst_i(w)$ for each $i \notin N$.*

As before, we must specify what the operation $b \oplus_N P$ amounts to:

Definition 5 $\oplus$ *is an* update function *relative an n-agent frame $\mathcal{F}$ and similarity function sim if and only if for any $P \in W_P$, any $b \in \mathcal{B}_i$ and any set of agents N (such that $i \in N$):*

$$\mathcal{W}(b \oplus_N P) = \{\oplus(w, P, N) | w \in sim_b(P)\}$$

The fixed point constraint now not only serves to deal with the agent's higher order beliefs, but also the fact that the other agents have changed their beliefs. Say that the original belief state of agent 1 contained only the world $w = \langle u, b_1, b_2, b_3 \rangle$ and say that agent 1 believes that agents 1 and 2 both update with P. The new belief state of agent 1 will then support the worlds $\{\langle \text{wst}(w'), \text{bst}(w') \oplus_{1,2} P, \text{bst}(w') \oplus_{1,2}, b_3 \rangle | w' \in \text{sim}_w(P)\}$. That is, the first-order factual beliefs of agent 1 will be the image of w under P and the first-order beliefs about agent 2 and 3 will be sensitive to the fact that the former has also updated with P while the latter hasn't.

Definition 6 *An* update *n-agent frame $\mathcal{F}_{\mathcal{U}}$ is an n-agent frame $\mathcal{F}$ extended with a similarity function sim and an update function $\oplus$.*

6 The logic

Define a formal language containing the following ingredients. $p, q, r, \ldots$ is a countable set of atoms. $\wedge, \neg, \vee, \rightarrow$ and $\leftrightarrow$ are the standard classical connectives. $\Rightarrow$ is the counter-factual conditional. $\oplus_N$ is the update function (here N is any finite set of natural numbers, bounded above by some number n, the number of agents). The *well formed formulas* (wff(s)) of the language are defined in standard fashion: the atoms are wffs; for any wffs A and B: $\neg A$, $A \wedge B$, $A \vee B$, $A \rightarrow B$, $A \leftrightarrow B$, $A \Rightarrow B$ are wffs; for any wffs A and B and finite set of natural numbers N (bounded above by n): $[\oplus A_N]B$ is a wff.

6.1 Semantics

The semantics will be given in terms of an *update model* which is an update n-agent frame extended with a valuation function V assigning a set of possible worlds to each atom p. The relation of truth in a world, $w \models A$, is defined inductively (and $\|A\|$ denotes the set of worlds $\{w \in W | w \models A\}$):

1. $w \models p$ if and only if $\text{wst}(w) \in V(p)$.
2. $w \models A \wedge B$ if and only if $w \models A$ and $w \models B$.
3. $w \models \neg A$ if and only if $w \not\models A$, and so on for the other boolean connectives.
4. $w \models \text{Bel}_i A$ if and only if for every $w' \in \mathcal{W}(\text{bst}_i(w)) : w' \models A$.
5. $w \models A \Rightarrow B$ if and only if for every $w' \in \text{sim}_w(\|A\|)$: $w' \models B$.
6. $w \models [\oplus_N A]B$ if and only if $\oplus(w, \|A\|, N) \models B$.

6.2 Axioms and rules

In addition to the standard set of axioms governing the classical connectives, we have three groups of axioms governing Bel, $\Rightarrow$ and $\oplus$, respectively.[2]

(NBel) $A/\mathrm{Bel}_i A$.

(KBel) $\mathrm{Bel}_i(A \rightarrow B) \rightarrow (\mathrm{Bel}_i A \rightarrow \mathrm{Bel}_i B)$.

(4) $\mathrm{Bel}_i A \rightarrow \mathrm{Bel}_i \mathrm{Bel}_i A$.

(5) $\neg \mathrm{Bel}_i A \rightarrow \mathrm{Bel}_i \neg \mathrm{Bel}_i A$.

The axioms for the counter-factual conditional $\Rightarrow$ reflect the fact that very weak conditions have been placed on the similarity function sim:

(N$\Rightarrow$) $A \rightarrow B/A \Rightarrow B$.

(K$\Rightarrow$) $(C \Rightarrow (A \rightarrow B)) \rightarrow ((C \Rightarrow A) \rightarrow (C \Rightarrow B))$.

(E$\Rightarrow$) $A \leftrightarrow B/(A \Rightarrow C) \leftrightarrow (B \Rightarrow C)$.

(MP$\Rightarrow$) $(A \wedge (A \Rightarrow B)) \rightarrow B$.

Finally the axioms for $\oplus$:

(N$\oplus$) $A/[\oplus_N B]A$.

(K$\oplus$) $[\oplus_N C](A \rightarrow B) \rightarrow ([\oplus_N C]A \rightarrow [\oplus_N C]B)$.

(E) $A \leftrightarrow B/[\oplus_N A]C \leftrightarrow [\oplus_N B]C$.

(F) $[\oplus_N A]\neg B \leftrightarrow \neg[\oplus_N A]B$.

(IND1) $p \leftrightarrow [\oplus_N A]p$, where p is a propositional letter.

(IND2) $\mathrm{Bel}_i B \leftrightarrow [\oplus_N A]\mathrm{Bel}_i B$, when $i \notin N$.

(GR) $[\oplus_N A]\mathrm{Bel}_i B \leftrightarrow \mathrm{Bel}_i(A \Rightarrow [\oplus_N A]B)$, when $i \in N$.

Theorem 1 *The axioms are sound with respect to the semantics.*

Proof: (GR) We have $w \models [\oplus_N A]\mathrm{Bel}_i B$ if and only if $\oplus(w, \|A\|, N) \models \mathrm{Bel}_i B$ if and only if $\mathcal{W}(\mathrm{bst}_i(w) \oplus_N \|A\|) \subseteq \|B\|$ (for $\mathrm{bst}_i(\oplus(w, \|A\|, N)) = \mathrm{bst}_i(w) \oplus_N \|A\|$) if and only if for every $w' \in \mathrm{sim}_{\mathrm{bst}_i(w)}(\|A\|)$: $\oplus(w', \|A\|, N) \in \|B\|$ (for $\mathcal{W}(\mathrm{bst}_i(w) \oplus_N \|A\|) = \bigcup\{\oplus(w', \|A\|, N) | w' \in \mathrm{sim}_{\mathrm{bst}_i(w)}(\|A\|)\}$) if and only if for every $w' \in \mathrm{sim}_{\mathrm{bst}_i(w)}(\|A\|)$: $w' \models [\oplus_N A]B$ if and only if $w \models \mathrm{Bel}_i(A \Rightarrow [\oplus_N A]B)$.

□

[2] The notation A/B is to be read: if A is a theorem, then so is B.

Theorem 2 *The axioms are complete with respect to the semantics.*

Proof: See appendix.

□

Most of the axioms are self-explanatory. The interesting axiom is (GR) (Generalized Ramsey).[3] It is *nearly* the axiom:

$$[\oplus_N A]\text{Bel}_i B \leftrightarrow \text{Bel}_i(A \Rightarrow B), \text{ when } i \in N.$$

That is, it comes close to stating that the agent will believe B after updating with A, if and only if the agent believes that if A were the case, then B would be the case. Such an axiom, however, would miss two points that (GR) captures. First, it is quite possible for an agent to believe: "even if A had been true, I would not have known that A was true". So even if one supposes that A were true, it doesn't follow that one then believes that A. By contrast a result of updating with A is that A is believed true. Second, the alternative axiom would miss the fact that the other agents in N have also updated with A.

So (GR) should be read: an agent will believe B after N-updating with A, if and only if the agent believes that if A were the case, then if the agents N-updated with this fact, B would be the case. What is interesting is that the axiom (GR) is nearly identical to the axiom with the same name presented in (Gerbrandy and Groeneveld, 1997) (and used in (Cantwell, 2005)) for the logic of common announcements. The one difference (which is of course very important) is that to get the logic of common announcements one replaces the counter-factual conditional with the material implication:

$$[\oplus_N A]\text{Bel}_i B \leftrightarrow \text{Bel}_i(A \rightarrow [\oplus_N A]B), \text{ when } i \in N.$$

As the remaining axioms are precisely those needed for common announcements, it is clear that the two operations are structurally very similar.

7 Concluding remarks

The model presented is very general in the sense that a group of agents can jointly update with any kind of proposition: factual propositions, *counter*-factual conditionals, propositions that make claims about what would happen if an agent were to update with some proposition A (that is, the agent's can N-update with propositions like $(A \Rightarrow B) \wedge [\oplus_M \text{Bel}_j A]\text{Bel}_k B$). When modeling a more specific application, like a particular game, it would be natural to restrict the expressiveness of the language (for instance, restricting the propositions that can generate an update to propositions like "player i moved pawn on E2 to E4"). Furthermore, the similarity function, which here has been left completely unanalyzed, could then be given a more determinate interpretation.

[3](Baltag and Moss, 2004) provides an axiom that is similar to (GR), although in a different kind of framework.

Appendix: Proof of Theorem 2

The completeness proof is based on standard methods of modal logic. Define:

- U is the set of maximal consistent sets of boolean formulae (formulas containing no doxastic and no dynamic operators).
- $V(p) = \{u \in U | p \in u\}$ (for any propositional letter p.
- $\mathcal{B}_i$ is the set of sets b of sentences such that for some maximal consistent set Γ: $b = \{\text{Bel}_i A | \text{Bel}_i A \in \Gamma\}$.

Given U and $\mathcal{B}$ the set W and the functions bst and wst can be defined.

- $\|A\| = \{w_\Sigma | \Sigma \text{ is a maximal consistent set and } A \in \Sigma\}$.

For any world $w = \langle u, b_1, \ldots, b_n \rangle$ let $\text{base}(w) = u \cup b_1 \cup \cdots \cup b_n$. Define:

- $\mathcal{W}(b_i) = \{w \in W | \forall C : \text{if Bel}_i C \in b_i, \text{ then } \{C\} \cup \text{base}(w) \text{ is consistent}\}$.
- For any $b \in \mathcal{B}_i$ and $i \in N$: $b \oplus_N \|A\| = \{\text{Bel}_i B | \text{Bel}_i [\oplus_N A] \text{Bel}_i B \in b\}$.
- $\text{sim}_w(\|A\|) = \{w' \in W | \forall B : \text{ if } A \Rightarrow B \in w, \text{ then } B \in w'\}$

For any maximal consistent set Σ, let $u_\Sigma = \{A \in \Sigma | A \text{ boolean}\}$ and $b_{i,\Sigma} = \{\text{Bel}_i A | \text{Bel}_i A \in \Sigma\}$. Let $w_\Sigma = \langle u_\Sigma, b_{1,\Sigma}, \cdots, b_{n,\Sigma} \rangle$. Clearly $w_\Sigma \in W$.

Lemma 1 *Let Σ be any maximal consistent set of sentences and A be any sentence. Then $\Gamma = \{B | [\oplus_N A] B \in \Sigma\}$ is a maximal consistent set such that $w_\Gamma = \oplus(w_\Sigma, N, \|A\|)$.*

Lemma 2

1. *If $w \in sim_b(\|A\|)$, then $\oplus(w, N, \|A\|) \in \mathcal{W}(b \oplus_N \|A\|)$.*
2. *If $w \in \mathcal{W}(b \oplus_N \|A\|)$, then for some $w' \in sim_b(\|A\|) : w = \oplus(w', N, \|A\|)$.*

Proof: (1) Assume that $w \in \text{sim}_b(\|A\|)$. There is some maximal consistent set Σ such that $w_\Sigma = w$. Let $\Gamma = \{B | [\oplus_N A] B \in \Sigma\}$. By Lemma 1 $w_\Gamma = \oplus(w, N, \|A\|)$. Assume for reductio that there is some $C \in \Gamma$ such that $\text{Bel}_i \neg C \in b \oplus_N \|A\|$. As $C \in \Gamma$, $[\oplus_N A] C \in \Sigma$. We also have (by construction) $\text{Bel}_i [\oplus_N A] \text{Bel}_i \neg C \in b$. By (GR) $\text{Bel}_i \text{Bel}_i (A \Rightarrow [\oplus_N A] \neg C) \in b$ and so $\text{Bel}_i (A \Rightarrow [\oplus_N A] \neg C) \in b$ and so $A \Rightarrow [\oplus_N A] \neg C \in \Sigma$. As $w_\Sigma \in \text{sim}_b(\|A\|)$, $A \in \Sigma$ so by (MP) $[\oplus_N A] \neg C \in \Sigma$. But then $\neg C \in \Gamma$ which would make Γ inconsistent.

So for all C such that $\text{Bel}_i C \in b \oplus_N \|A\|$, $C \in \Gamma$. But then $w_\Gamma \in \mathcal{W}(b \oplus_N \|A\|)$.

(2) Assume that $w \in \mathcal{W}(b \oplus_N \|A\|)$. The aim is to show that there is a $w' \in \text{sim}_b(\|A\|)$ such that $w = \oplus(w', \|A\|, N)$.

Let $X = \text{wst}(w)$, $Y = \bigcup \{\text{bst}_j(w) | j \notin N\}$ and $Z = \bigcup \{\text{Bel}_j B | j \in N \text{ and } \text{Bel}_i (A \Rightarrow \text{Bel}_j B) \in b\}$.

Assume that there is some element $B \in X$ and some $\text{Bel}_{j_1} C_1, \ldots, \text{Bel}_{j_m} C_m \in Y$ (let $E = B \wedge \text{Bel}_{j_1} C_1 \wedge \cdots \wedge \text{Bel}_{j_m} C_m$) and some $\text{Bel}_{k_1} D_1, \ldots, \text{Bel}_{k_l} D_l \in Z$ (let $F = \text{Bel}_{k_1} D_1 \wedge \cdots \wedge \text{Bel}_{k_l} D_l$) such that $\text{Bel}_i(A \Rightarrow \neg(E \wedge F)) \in b$.

We know that $\text{Bel}_i \neg E \notin b \oplus_N \|A\|$. So $\text{Bel}_i \neg \text{Bel}_i \neg E \in b \oplus_N \|A\|$. Thus $\text{Bel}_i [\oplus_N A] \text{Bel}_i \neg \text{Bel}_i \neg E \in b$. So, by introspection, $\text{Bel}_i [\oplus_N A] \neg \text{Bel}_i \neg E \in b$. By (F), $\text{Bel}_i \neg [\oplus_N A] \text{Bel}_i \neg E \in b$. By (GR) $\text{Bel}_i \neg (A \Rightarrow [\oplus_N A] \neg E) \in b$.

By (IND1) and (IND2), $\text{Bel}_i \neg (A \Rightarrow \neg E) \in b$. But then $\text{Bel}_i \neg (A \Rightarrow F) \in b$, however, by construction, $\text{Bel}_i (A \Rightarrow F) \in b$, so then we have $\text{Bel}_i (p \wedge \neg p) \in b$. But then by (IND1) $\text{Bel}_i (p \wedge \neg p) \in b \oplus_N \|A\|$ which means that $\mathcal{W}(b \oplus_N \|A\|)$ is empty, contrary to assumption. So $\text{Bel}_i (A \Rightarrow \neg (E \wedge F)) \notin b$.

Thus for any B that is a consequence of $X \cup Y \cup Z$, $\text{Bel}_i (A \Rightarrow \neg B) \notin b$. Hence every consequence of $X \cup Y \cup Z$ contradicts A, which means that $X \cup Y \cup Z \cup \{A\}$ is a consistent set and can be extended to a maximal consistent set Σ.

Furthemore $A \in \Sigma$. So $\{B | A \Rightarrow B \in b\} \subseteq \Sigma$, but then $w_\Sigma \in \text{sim}_b(\|A\|)$. Furthermore, for each $i \notin N$ we have $\text{bst}_i(w_\Sigma) = \text{bst}_i(w)$ and for each $i \in N$ we have $\text{bst}_i(w) = \text{bst}_i(w_\Sigma) \oplus \|A\|$. We also have $\text{wst}(w_\Sigma) = \text{wst}(w)$. So $w = \oplus(w_\Sigma, \|A\|, N)$.

□

We now know that $\langle W_P, U, \mathcal{B}_i, \gamma, \oplus, V \rangle$ is an n-agent dynamic model.

Lemma 3 *For any maximal consistent set Σ and any formula A:*

$$w_\Sigma \models A \text{ if and only if } A \in \Sigma$$

Proof: By induction over the length of A. The interesting case is when A is of the form $[\oplus_N B]C$. So (i) $w_\Sigma \models [\oplus_N B]C$ if and only if (ii) $\oplus(w_\Sigma, \|B\|, N) \models C$. By Lemma 1 $w_\Gamma = \oplus(w_\Sigma, \|B\|, N)$ where $\Gamma = \{B | [\oplus_N A] B \in \Sigma\}$. By the induction hypothesis $w_\Gamma \models C$ if and only if $C \in \Gamma$. So (ii) holds if and only if (iii) $C \in \Gamma$ if and only if (iv) $[\oplus_N A] C \in \Sigma$.

□

So for any consistent set of sentences Δ there is a model and a world in the model that satisfies Δ. This concludes the proof.

References

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

Baltag, A. and Moss, L. S. (2004). Logics for epistemic programs. *Knowledge, Rationality and Action*, pages 1–60.

Baltag, A., Moss, L. S., and Solecki, S. (1999). The logic of public announcements, common knowledge, and private suspicions. *Technical report CWI.*

Cantwell, J. (2005). A formal-model of multi-agent belief interaction. *Journal of Logic, Language and Information.* To appear.

Fagin, R., Halpern, J. Y., Moses, Y., , and Vardi, M. Y. (1995). *Reasoning About Knowledge*. MIT Press.

Gerbrandy, J. (1999). *Bisimulations on planet Kripke*. Ph.D dissertation, University of Amsterdam, ILLC dissertation series, Amsterdam.

Gerbrandy, J. and Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language and Information*, 6:147–169.

Joyce, J. (1999). *The foundations of causal decision theory*. Cambridge University Press, Cambridge.

Katsuno, H. and Mendelzon, A. (1992). On the difference between updating a knowledge base and revising it. In Gärdenfors, P., editor, *Belief Revision*, pages 183–203. Cambridge University Press, Cambridge.

Kooi, B. P. (2003). *Knowledge, Chance and Change*. PhD dissertation, University of Groningen, ILLC Dissertation Services, Amsterdam.

Levi, I. (1996). *For the sake of the argument*. Cambridge University Press, Cambridge.

Lewis, D. (1973). *Counterfactuals*. Harvard University Press, Cambridge.

Moss, L. S. (1999). From hypersets to Kripke models in logics of announcements. In et al., J. G., editor, *JFAK. Essay dedicated to Johan van Benthem on the occasion of his fiftieth birthday*. Amsterdam University Press.

Plaza, J. (1989). Logics of public communications. In *Proceedings, 4th International Symposium on Methodologies for Intelligent Systems*.

Segerberg, K. (1996). A general framework for the logic of theory change. *Bulletin of the Section of Logic*, 25:2–8.