Department of Computer Science,
University of Otago

Technical Report OUCS-2006-11

The logic of ontic and epistemic change

Authors:

Hans van Ditmarsch
Department of Computer Science, University of Otago

Barteld Kooi
Department of Philosophy, University of Groningen

Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

http://www.cs.otago.ac.nz/research/techreports.html

The logic of ontic and epistemic change

Hans van Ditmarsch* Barteld Kooi†

October 14, 2006

Abstract

We propose an epistemic logic incorporating dynamic operators to describe information changing events, both informative actions, where agents become more informed about the non-changing state of the world, as ontic changes, wherein the world and the facts describing it change themselves as well. A complete axiomatisation is provided. There are some independent semantic results, for example that assignments in programs can be restricted to those to false or true only. We apply the logic to model systems for card deals and to model protocols.

1 Introduction

In dynamic epistemic logics [27] one does not merely describe the static (knowledge and) beliefs of agents but also dynamic features: how does belief change as a result of events taking place. The main focus of such logics has been change of *only* belief, whereas the facts describing the world remain the same. Change of belief is known as *epistemic* change, or information change. Alternatively or additionally, one can model change of facts, and the resulting consequences of such factual changes for the beliefs of the agents. Change of facts is also known as *ontic* change (change of the real world, so to speak). From the somewhat different perspective of the areas of artificial intelligence and belief revision, epistemic and ontic change correspond to, respectively, revising [1] and updating [12] beliefs. Change, revision, action, ontic,... there is much overlapping terminology around. In this contribution we use 'event' (and its formal counterpart 'update'), to denote *any* sort of change, both epistemic and ontic. Let us begin by a simple example involving various events.

Example 1 Given are two players Anne and Bill. Anne shakes a cup containing a single coin and deposits the cup upside down on the table (there are no opportunities for cheating). Heads or tails? Let atomic proposition p stand for 'the coin lands heads'. Initially, we have a situation wherein both Anne (a) and Bill (b) are uncertain about p. This is represented by a two-state epistemic model with domain $\{1, 0\}$, with universal access for a and b, and with $V(p) = \{1\}$. A player may observe whether the coin is heads or tails, and/or flip the coin, and with or without the other player noticing that. Four detailed events are as follows.

1. Anne lifts the cup and looks at the coin. Bill observes this but is not able to see the coin. (All the previous is common knowledge.)
2. Anne lifts the cup and looks at the coin without Bill noticing that.
3. Anne lifts the cup, looks at the coin, and ensures that it is tails (by some sleight of hand). Bill observes Anne looking but is not able to see the coin; he considers it possible that Anne may have flipped the coin to tails (and this is common knowledge).

*University of Otago, New Zealand, hans@cs.otago.ac.nz
†University of Groningen, Netherlands, B.P.Kooi@rug.nl

4. Bill flips the coin (without seeing it). Anne considers that possible (and this is common knowledge).

Events 1, 3, and 4 are all public in the sense that the actual state of the world is considered possible by both agents, and that both agents know that, and know that they know that, etc.; whereas event 2 is private: Bill is unaware of the action. Events 3 and 4 involve ontic change, whereas events 1 and 2 only involve epistemic change. We will later model these examples in the logical language. $\dashv$

Dynamic modal operators for ontic change, in addition to similar operators for epistemic change, have been suggested in various recent publications. As far as we know it was first mentioned by Baltag, Moss, and Solecki as a possible extension to their *action model logic* (for epistemic change), in [4]. This was by example only and without a language or a logic. A precise quotation of all these authors say on ontic change may be in order:

> *Our second extension concerns the move from actions as we have been working them to actions which change the truth values of atomic sentences. If we make this move, then the axiom of Atomic Permanence*[1] *is no longer sound. However, it is easy to formulate the relevant axioms. For example, if we have an action* α *which effects the change* $p := p \wedge \neg q$*, then we would take an axiom* $[\alpha]p \leftrightarrow (PRE(\alpha) \rightarrow p \wedge \neg q)$*. Having made these changes, all the rest of the work we have done goes through. In this way, we get a completeness theorem for this logic.* [4, p.24]

The logic that we present here is a realization of their proposal, although it is unclear to us if the authors correctly assessed the required theoretical and modelling effort when stating "all the rest (...) goes through". In a later proposal by Baltag [2] a fact changing action $flipP$ is proposed that changes ('flips') the truth value of an atom P, with accompanying axioms (for the proper correspondent action α resembling a single-pointed action model) $[\alpha]p \leftrightarrow (\mathsf{pre}(\alpha) \rightarrow \neg p)$ if "p changes value (flips) in α", and otherwise $[\alpha]p \leftrightarrow (\mathsf{pre}(\alpha) \rightarrow p)$ [2, p.29]. The approach is restricted to ontic change where the truth value of atoms flips. In the concluding section of [2], the author defers the relation of this proposal to a general logic of ontic and epistemic change to the future.

More recently, in a MAS application-driven line of research [26, 25] assignments are added to the relational action language of [22] but without providing an axiomatization. In this setting only change of *knowledge* is modelled and not change of belief, i.e., such actions describe transformation of $S5$ models only, such as Kripke models corresponding to interpreted systems. A line of research culminating in [21] also combines epistemic and ontic change. This should not be seen as dynamic epistemics (epistemic logic, with dynamic operators corresponding to the execution of epistemic programs) but rather as epistemic dynamics (dynamic logic, in an epistemic interpretation), namely PDL extended with dynamic operators for action models. Our treatment of postconditions, also called substitutions, is the same as in [21]; unlike them, we focus on specific semantic results and applications, and, as said, we use a dynamic epistemic 'dialect'. On the whole, their treatment is much more general. A recent contribution on combining *public* ontic and epistemic change, including detailed expressivity results for different combinations of static and dynamic modalities, is found in [13]; our work uses a similar approach to ontic events but describes more complex than public events. Finally, a general dynamic modal logic is presented in [17], where ontic changes are also studied. The semantics of this logic uses tree-like structures for its semantics, and fixed points are introduced in the language to be able to reason about updates.

An independent recent line of investigation combining epistemic with ontic change arises from the belief revision community. Modelling *belief revision*, i.e., epistemic change, by dynamic operators is a old idea going back to Van Benthem [19]. (In fact, this is one of the two original publications—together with [16]—that starts the area of dynamic epistemic logic. For an overview of such matters see [5, 24, 27].) But modelling ontic change—known as *belief update* [12]—in a similar, dynamic modal, way, including its interaction with epistemic change, is only a recent

[1] I.e.: $[\alpha]p \leftrightarrow (PRE(\alpha) \rightarrow p)$ [4, p.15]

focus of ongoing research by Herzig and collaborators [11, 10]. Their work sees the execution of an event as so-called *progression* of information, and reasoning from a final information state to a sequence of events realizing it as *regression*—the last obviously relates to planning. The focus of progression and regression is the change of the *theory* describing the information state, i.e. the set of all true, or believed, formulas. See as well [14], a setting also involving degrees of belief.

In this contribution we combine ontic and epistemic change, as in [13], but for the full generality of arbitrarily complex events, such as exchange of cards among subgroups of the public, and other events with a 'private' (as opposed to public) character. A novel feature, also present in [13, 21] but not in [4], is to define assignments / postconditions (expressing ontic change) concurrently for a *finite* subset of atomic propositional letters. Various semantic results, probably just as applicable to other 'dialects' of action languages, form a major part of our contribution. These results are that: for all finite models and for all consistent formulas we can construct an event that 'realizes' the formula, i.e., makes the formula true; and that: arbitrary events (with assignments of form $p := \varphi$ for φ in the language) correspond to events with assignments to true and false only; and also that: arbitrary events correspond to events with assignments for a single atom only. We provide a full axiomatization of the logic. This had best be seen as an alternative presentation of the completeness proof in [4], including extensions to handle assignments. We then apply the logic to model the dynamics of two different multi-agent systems: various actions involving seeing, showing, and trading cards; and a protocol to solve the 'one hundred prisoners and a lightbulb' problem.[2] We deem the applications a substantial part of our contribution.

In Section 2 we define structures representing knowledge (and belief), and similar structures representing events combining epistemic and ontic change. In Section 3 we define the logic for those events. Subsection 3.1 provides the axiomatization of the logic, and is followed by subsections with semantic observations of specific interest, as mentioned above. Details on the completeness proof in the appendix. The applications are in Section 5.

2 Updates formalizing epistemic and ontic change

The models which adequately present an information state in a multi-agent environment are Kripke models from epistemic logic. The set of states together with the accessibility relations represent the information the agents have. If one state s has access to another state t for an agent a, this means that, if the actual situation is s, then according to a's information it is possible that t is the actual situation.

Definition 2 (Epistemic model) Let a finite set of agents A and a countable set of propositional variables P be given. An epistemic model is a triple $M = (S, R, V)$ such that

- S is a non-empty set of possible states,
- $R : A \to \wp(S \times S)$ assigns an accessibility relation to each agent a,
- $V : P \to \wp(S)$ assigns a set of states to each propositional variable.

A pair (S, R) is called an *epistemic frame*. A pair (M, s), with $s \in S$, is called an *epistemic state*.[3] ⊣

[2]In Section 5.2 we pose and then straightaway *solve* the riddle. For those who do not know it but want to ponder it first: *A group of 100 prisoners, all together in the prison dining area, are told that they will be all put in isolation cells and then will be interrogated one by one in a room containing a light with an on/off switch. The prisoners may communicate with one another by toggling the light-switch (and that the only way in which they can communicate). The light is initially switched off. There is no fixed order of interrogation, or interval between interrogations, and the same prisoner may be interrogated again at any stage. When interrogated, a prisoner can either do nothing, or toggle the light-switch, or announce that all prisoners have been interrogated. If that announcement is true, the prisoners will (all) be set free, but if it is false, they will all be executed. While still in the dining room, and before the prisoners go to their isolation cells (forever), can the prisoners agree on a protocol that will set them free?*

[3]'Epistemic state' in our sense is not to be confused with the set of worlds/states indistinguishable for a given agent from a given state, i.e., $R(a)(s) = \{t \mid (s,t) \in R(a)\}$. This set $R(a)(s)$ is also commonly called an epistemic state. We call that the set of reachable states, or, in the case of $KD45$ or $S5$-agents, the epistemic class.

An epistemic model represents the information of the agents. *Information change* should therefore be modelled as changes of such a model. There are three variables. One can change the set of states, the accessibility relations and the valuation. It may be difficult to find the exact change of these variables that matches a certain description of an information changing event. It is often easier to think of such an event separately. One can model an information changing event in the same way as an information state, namely as some kind of Kripke model: there are various possible events, which the agents may not be able to distinguish. This is the domain of the model. Rather than a valuation, a precondition captures the conditions under which such events may occur. Such a Kripke model for events is called an *update model*.

Definition 3 (Update model) An update model for a finite set of agents A and a language $\mathscr{L}$ is a quadruple $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$ where

- E is a finite non-empty set of events,
- $\mathsf{R} : A \to \wp(\mathsf{E} \times \mathsf{E})$ assigns an accessibility relation to each agent,
- $\mathsf{pre} : \mathsf{E} \to \mathscr{L}$ assigns a precondition to each event,
- $\mathsf{post} : \mathsf{E} \to (P \rightharpoonup \mathscr{L})$ assigns a partial substitution with a finite domain to each event. This is called the *postcondition* of that event.

A pair (U, e) with a distinguished actual event $\mathsf{e} \in \mathsf{E}$ is called an *update*. ⊣

Instead of

$$\mathsf{pre}(\mathsf{e}) = \varphi \textit{ and } \mathsf{post}(\mathsf{e})(p_1) = \psi_1, \textit{ ... and } \mathsf{post}(\mathsf{e})(p_n) = \psi_n$$

we also write

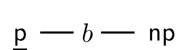
$$\textit{given event } \mathsf{e}, \textit{ if } \varphi, \textit{ then } p_1 := \psi_1, \textit{ ..., and } p_n := \psi_n.$$

The event e with $\mathsf{pre}(\mathsf{e}) = \top$ and $\mathsf{post}(\mathsf{e}) = \emptyset$ we name skip. This stands for 'nothing happens' (except a 'tick of the clock'). An update with a singleton domain, accessible to all agents, and precondition $\top$, is a *(public) assignment*. An update with a singleton domain, accessible to all agents, and empty postconditions, is a *(public) announcement*.

The finite domain restriction on postconditions is needed because we will, in Section 3, propose updates (U, e) as primitives in the logical language (similar to adding automata to PDL). Without that restriction updates are infinitary constructs if P is infinite. Still, it is occasionally more convenient to present the postcondition function for a given event as a total function. We therefore propose to abuse the language and also write $\mathsf{post}(\mathsf{e})(p) = p$ if $p \notin \mathrm{dom}(\mathsf{post})$, thus informally treating $\mathsf{post}(\mathsf{e})$ as a total function.

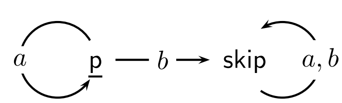
Example 4 The events of Example 1 on page 1 can be visualized as the following updates. Assume reflexivity and symmetry of access for a and b in all visualizations except for item 2 wherein all access is explicit. The coin is heads when p is true. The actual state is underlined.

1. Anne lifts the cup and looks at the coin. Bill observes this but is not able to see the coin.

   Here, $\mathsf{pre}(\mathsf{p}) = p$, $\mathsf{post}(\mathsf{p}) = \emptyset$, $\mathsf{pre}(\mathsf{np}) = \neg p$, $\mathsf{post}(\mathsf{np}) = \emptyset$.

2. Anne lifts the cup and looks at the coin without Bill noticing that.

   Event p is as in the previous item and skip is as above.

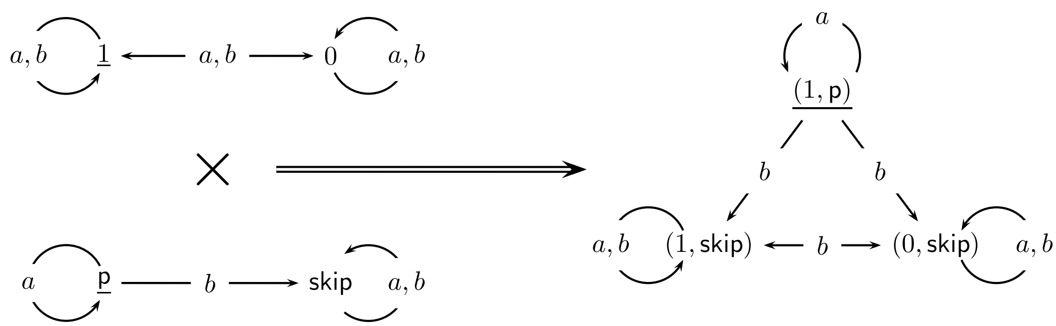
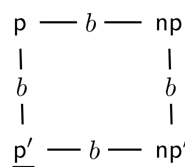


Figure 1: In an epistemic state where Anne and Bill are uncertain about the truth of p (heads or tails), and wherein p is true, Anne looks at the coin without Bill noticing it. The actual states are underlined.

3. Anne lifts the cup, looks at the coin, and ensures that it is tails. Bill observes Anne, and considers it possible that she does nothing and also that Anne ensures tails.



   Events p and np are as before, whereas $\mathsf{pre}(\mathsf{p}') = p$, $\mathsf{post}(\mathsf{p}')(p) = \bot$, $\mathsf{pre}(\mathsf{np}') = \neg p$, $\mathsf{post}(\mathsf{np}')(p) = \bot$. Note that the event where Anne ensures *an* outcome (but not necessary 'tails') would correspond to yet another update, namely one consisting of six different events.

4. Bill succeeds in flipping the coin under the cup (without seeing it) – Anne considers that possible.

   $\underline{\mathsf{n}}$ — a — skip

   Where $\mathsf{pre}(\mathsf{n}) = \top$, $\mathsf{post}(\mathsf{n})(p) = \neg p$, and skip as before. ⊣

Now that we have a way of modelling updates, we want to determine what the effects of these information changing events are on an information state. This is given by the following definition—Figure 1 gives an example of such update execution.

Definition 5 (Execution) Given are an epistemic model $M = (S, R, V)$, a state $s \in S$, an update model $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$ for a language $\mathscr{L}$ that can be interpreted in M, and an event $\mathsf{e} \in \mathsf{E}$ with $(M, s) \models \mathsf{pre}(\mathsf{e})$. The result of executing (U, e) in (M, s) is the model $(M \otimes \mathsf{U}, (s, \mathsf{e})) = ((S', R', V'), (s, \mathsf{e}))$ where

- $S' = \{(t, \mathsf{f}) \mid (M, t) \models \mathsf{pre}(\mathsf{f})\}$,
- $R'(a) = \{((t, \mathsf{f}), (u, \mathsf{g})) \mid (t, u) \in R(a) \text{ and } (\mathsf{f}, \mathsf{g}) \in \mathsf{R}(a)\}$,
- $V'(p) = \{(t, \mathsf{f}) \mid (M, t) \models \mathsf{post}(\mathsf{f})(p)\}$. ⊣

3 The logic of ontic and epistemic change

The models from the previous section can be used to define a logic for reasoning about information change. An update is associated with a dynamic operator in a modal language, based on epistemic logic. Unlike above, where they introduced relative to a *given* logical language, the updates are now part of the language: an update (U, e) is an inductive construct of type α that should be seen as built from simpler constructs of type φ, namely the preconditions and postconditions for the events of which the update consists. Note that only for only a *finite* number of atoms postconditions are given, otherwise updates would be infinitary language constructs. Because updates are part of the language, we also restrict ourselves to *finite* update models because they would otherwise still be infinitary.

Definition 6 (Language) Let a finite set of agents A and a countable set of propositional variables P be given. The language $\mathscr{L}$ is given by the following BNF:

$$\begin{array}{lcl} \varphi & ::= & p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\alpha]\varphi \\ \alpha & ::= & a \mid B^+ \mid (\mathsf{U}, \mathsf{e}) \end{array}$$

where U is a finite update model for A and $\mathscr{L}$. We use the usual abbreviations. $\dashv$

The semantics of this language is standard for epistemic logic and based on the product construction for the execution of update models from the previous section. Below, $R(B)^+$ is the transitive closure of the union of all accessibility relations $R(a)$ for agents $a \in B$.

Definition 7 (Semantics) Let a model (M, s) with $M = (S, R, V)$ be given. Let $a \in A$, $B \subseteq A$, and $\varphi, \psi \in \mathscr{L}$.

$$\begin{array}{lll} (M,s) \models p & \text{iff} & s \in V(p) \\ (M,s) \models \neg\varphi & \text{iff} & (M,s) \not\models \varphi \\ (M,s) \models \varphi \wedge \psi & \text{iff} & (M,s) \models \varphi \text{ and } (M,s) \models \psi \\ (M,s) \models [a]\varphi & \text{iff} & (M,t) \models \varphi \text{ for all } t \text{ such that } (s,t) \in R(a) \\ (M,s) \models [B^+]\varphi & \text{iff} & (M,t) \models \varphi \text{ for all } t \text{ such that } (s,t) \in R(B)^+ \\ (M,s) \models [\mathsf{U},\mathsf{e}]\varphi & \text{iff} & (M,s) \models \mathsf{pre}(\mathsf{e}) \text{ implies } (M \otimes \mathsf{U}, (s,\mathsf{e})) \models \varphi \end{array}$$

Definition 8 (Composition of update models) Let two update models $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$ and $\mathsf{U}' = (\mathsf{E}', \mathsf{R}', \mathsf{pre}', \mathsf{post}')$ and two events $\mathsf{e} \in \mathsf{E}$ and $\mathsf{e}' \in \mathsf{E}'$ be given. The composition $(\mathsf{U}, \mathsf{e})\ ;\ (\mathsf{U}', \mathsf{e}')$ of these update models is $(\mathsf{U}'', \mathsf{e}'')$ where $\mathsf{U}'' = (\mathsf{E}'', \mathsf{R}'', \mathsf{pre}'', \mathsf{post}'')$ is defined as follows

- $\mathsf{E}'' = \mathsf{E} \times \mathsf{E}'$,
- $\mathsf{R}''(a) = \{((\mathsf{f}, \mathsf{f}'), (\mathsf{g}, \mathsf{g}')) \mid (\mathsf{f}, \mathsf{g}) \in \mathsf{R}(a) \text{ and } (\mathsf{f}', \mathsf{g}') \in \mathsf{R}'(a)\}$,
- $\mathsf{pre}''(\mathsf{f}, \mathsf{f}') = \mathsf{pre}(\mathsf{f}) \wedge [\mathsf{U}, \mathsf{f}]\mathsf{pre}'(\mathsf{f}')$,
- $\mathrm{dom}(\mathsf{post}''(\mathsf{f}, \mathsf{f}')) = \mathrm{dom}(\mathsf{post}(\mathsf{f})) \cup \mathrm{dom}(\mathsf{post}'(\mathsf{f}'))$ and if $p \in \mathrm{dom}(\mathsf{post}''(\mathsf{f}, \mathsf{f}'))$, then

$$\mathsf{post}''(\mathsf{f}, \mathsf{f}')(p) = \begin{cases} \mathsf{post}(\mathsf{f})(p) & \text{if } p \notin \mathrm{dom}(\mathsf{post}'(\mathsf{f}')), \\ [\mathsf{U}, \mathsf{f}]\mathsf{post}'(\mathsf{f}')(p) & \text{otherwise.} \end{cases}$$

The occurrence of $[\mathsf{U}, \mathsf{f}]\mathsf{post}'(\mathsf{f}')(p)$ in the final clause will become clear in the proof of the following proposition.

Proposition 9 $\models [\mathsf{U}, \mathsf{e}][\mathsf{U}', \mathsf{e}']\varphi \leftrightarrow [(\mathsf{U}, \mathsf{e})\ ;\ (\mathsf{U}', \mathsf{e}')]\varphi$ $\dashv$

Proof Let (M, t) be arbitrary. To show that $(M, t) \models [(\mathsf{U}, \mathsf{e})\ ;\ (\mathsf{U}', \mathsf{e}')]\varphi$ if and only if $(M, t) \models [\mathsf{U}, \mathsf{e}][\mathsf{U}', \mathsf{e}']\varphi$, it suffices to show that $M \otimes (\mathsf{U}\ ;\ \mathsf{U}')$ is isomorphic to $(M \otimes \mathsf{U}) \otimes \mathsf{U}')$.

Domain Let $(t, (\mathsf{e}, \mathsf{e}')) \in \mathrm{dom}(M \otimes (\mathsf{U}\ ;\ \mathsf{U}'))$. We then have that $(M, t) \models \mathsf{pre}''((\mathsf{e}, \mathsf{e}'))$, i.e., $M, t \models \mathsf{pre}(\mathsf{e}) \wedge [\mathsf{U}, \mathsf{e}]\mathsf{pre}'(\mathsf{e}')$, i.e., $M, t \models \mathsf{pre}(\mathsf{e})$ and $M, t \models [\mathsf{U}, \mathsf{e}]\mathsf{pre}'(\mathsf{e}')$. From $M, t \models \mathsf{pre}(\mathsf{e})$ follows that $(t, \mathsf{e}) \in \mathrm{dom}(M \otimes \mathsf{U})$, and from that and $M, t \models [\mathsf{U}, \mathsf{e}]\mathsf{pre}'(\mathsf{e}')$ follows that $((t, \mathsf{e}), \mathsf{e}') \in \mathrm{dom}((M \otimes \mathsf{U}) \otimes \mathsf{U}')$. The argument runs both ways.

all instantiations of propositional tautologies	
$[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$	distribution
From φ and $\varphi \to \psi$, infer ψ	modus ponens
From φ, infer $[\alpha]\varphi$	necessitation
$[\mathsf{U},\mathsf{e}]p \leftrightarrow (\mathsf{pre}(\mathsf{e}) \to \mathsf{post}_{\mathsf{e}}(p))$	update and atoms
$[\mathsf{U},\mathsf{e}]\neg\varphi \leftrightarrow (\mathsf{pre}(\mathsf{e}) \to \neg[\mathsf{U},\mathsf{e}]\varphi)$	update and negation
$[\mathsf{U},\mathsf{e}](\varphi \wedge \psi) \leftrightarrow ([\mathsf{U},\mathsf{e}]\varphi \wedge [\mathsf{U},\mathsf{e}]\psi)$	update and conjunction
$[\mathsf{U},\mathsf{e}][a]\varphi \leftrightarrow (\mathsf{pre}(\mathsf{e}) \to \bigwedge_{(\mathsf{e},\mathsf{f})\in\mathsf{R}(a)}[a][\mathsf{U},\mathsf{f}]\varphi)$	update and knowledge
$[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\varphi \leftrightarrow [(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')]\varphi$	update composition
$[B^+]\varphi \to [B](\varphi \wedge [B^+]\varphi)$	mix
$[B^+](\varphi \to [B]\varphi) \to ([B]\varphi \to [B^+]\varphi)$	induction axiom
Let (U,e) be an update model and let a set of formulas χ_{f} for every f such that $(\mathsf{e},\mathsf{f}) \in \mathsf{R}(B)^+$ be given. From $\chi_{\mathsf{f}} \to [\mathsf{U},\mathsf{f}]\varphi$ and $(\chi_{\mathsf{f}} \wedge \mathsf{pre}(\mathsf{f})) \to [a]\chi_{\mathsf{g}}$ for every $\mathsf{f} \in \mathsf{E}$, $a \in B$ and $(\mathsf{f},\mathsf{g}) \in \mathsf{R}(a)$, infer $\chi_{\mathsf{e}} \to [\mathsf{U},\mathsf{e}][B^+]\varphi$.	updates and common knowledge

Table 1: The proof system **UM**.

Accessibility $((t,(\mathsf{e},\mathsf{e}')),(t_1,(\mathsf{e}_1,\mathsf{e}'_1))) \in R(a)$ iff ($R(a)(t,t_1)$ and $R(a)(\mathsf{e},\mathsf{e}_1)$ and $R(a)(\mathsf{e}',\mathsf{e}'_1)$) iff $R(a)(((t,\mathsf{e}),\mathsf{e}'),((t_1,\mathsf{e}_1),\mathsf{e}'_1))$.

Valuation For the valuation of facts p in the domain of post'' we distinguish the cases ($p \in \mathrm{dom}(\mathsf{post}(\mathsf{e}))$ but $p \notin \mathrm{dom}(\mathsf{post}'(\mathsf{e}'))$) (i), and otherwise (ii). The valuation of a *i*-atom in a triple $(t,(\mathsf{e},\mathsf{e}'))$ is $\mathsf{post}(\mathsf{e})(p)$ according to the definition of updates composition; and the valuation of a *ii*-atom is $[\mathsf{U},\mathsf{e}]\mathsf{post}'(\mathsf{e}')$. Consider the corresponding triple $((t,\mathsf{e}),\mathsf{e}')$. The valuation of an *i*-atom in (t,e) is $\mathsf{post}(\mathsf{e})(p)$, and because p does not occur in $\mathrm{dom}(\mathsf{post}'(\mathsf{e}'))$ its value in the triple $((t,\mathsf{e}),\mathsf{e}')$ will remain the same. For a *ii*-atom, its final value will be determined by evaluating $\mathsf{post}'(\mathsf{e}')(p)$ in $((M \otimes \mathsf{U}),(t,\mathsf{e}))$. This corresponds to evaluating $[\mathsf{U},\mathsf{e}]\mathsf{post}'(\mathsf{e}')(p)$ in (M,t).

For examples we refer to the applications in Section 5.

3.1 Proof system and completeness

A proof system **UM** for the logic is given in Table 1. The proof system is a lot like the proof system for the logic of epistemic actions in [4]. There are two differences. The axiom 'atomic permanence' in [4]—$[\mathsf{U},\mathsf{e}]p \leftrightarrow (\mathsf{pre}(\mathsf{e}) \to p)$—is now instead an axiom expressing when atoms are *not* permanent, namely how the value of an atom can change, according to the postcondition for that atom:

$$[\mathsf{U},\mathsf{e}]p \leftrightarrow (\mathsf{pre}(\mathsf{e}) \to \mathsf{post}_{\mathsf{e}}(p)) \qquad \text{update and atoms}$$

The second difference is not apparent from Table 1. The axiom

$$[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\varphi \leftrightarrow [(\mathsf{U},\mathsf{e})\ ;\ (\mathsf{U}',\mathsf{e}')]\varphi \qquad \text{update composition}$$

that also occurs in [4] now instead uses Definition 8 to compute that composition, a more complex construction than the composition of action models with only preconditions. We find it remarkable that these are the only differences: the interaction between *postconditions* for an atom and the logical operators, *only* occurs in the axiom where that atom is mentioned, or implicitly, whereas the interaction between *preconditions* and the logical operators appears in several axioms and rules.

The proof system is sound and complete. The soundness of the 'update and atoms' axiom is evident. The soundness of the 'composition of update models' axiom was established in Proposition 9. The completeness proof is an inessential variant of the proof in [4]. We were therefore uncertain if it would be informative to the reader. We have chosen to include the full proof in the

Appendix on page 22. Our considerations to include it, were (*i*) to make explicit some technical differences between our proof (finite closure sets, and numerical complexity) and [4] (filtration, and lexicographic order); (*ii*) to let the reader determine exactly where postconditions play a role in the proof, and (*iii*) the regrettable circumstance that [4], although a publicly available report, has not been published in a journal: the semantics of action model logic has found its final form in [3], but the presentation of the completeness results and various expressivity results are still work in progress (personal communication). We continue with some details on points (*i*) and (*ii*) above.

(*i*) Our method to establish completeness builds a finite model satisfying a given formula, that consists of maximal consistent subsets of a Fischer-Ladner-like closure of that formula. Instead, [4] use a filtration technique on an (possibly) infinite model of maximal consistent sets. This is not a big difference. Baltag et al. [4, p.29] use a lexicographic path order wherein, in simple words, dynamic parts of the language ('updates') are more complex than the static parts of the language ('formulas'). Formulas in the language can be rewritten to certain 'normal forms' of lower complexity in this lexicographic path order. Instead, we use a direct numerical complexity on formulas in the finite closure, with essential clause $c([\mathsf{U},\mathsf{e}]\varphi) = (4+c(\mathsf{U}))\cdot c(\varphi)$, see Definition 26, on page 26. This complexity plays a role in Truth Lemma 28, on page 28. The same complexity is used in the textbook [27] to prove completeness of public announcement logic and action model logic, and the number 4 in $(4+c(\mathsf{U}))$ is the *minimum* natural number that achieves the result. This may be considered an interesting simplification compared to a completeness proof with lexicographic orders. But the *essential* contribution of [4] is of course the derivation rule 'updates and common knowledge' and its role in the proof: we copied the exact approach of [4] in that matter.

(*ii*) The postconditions play a role in the various parts of the completeness proof. The closure set $cl(\varphi)$ of a formula φ is the finite set of formulas that are 'relevant' for the construction of the canonical model for φ. Clearly, if φ contains a subformula of the form $[\mathsf{U},\mathsf{e}]\psi$ we expect the postconditions of events in the domain of U to reappear in closure $cl(\varphi)$ in some form. It appears in the basic clause of the closure definition: 'if $[\mathsf{U},\mathsf{e}]p \in cl(\varphi)$ then $(\mathsf{pre}(\mathsf{e}) \rightarrow \mathsf{post}(\mathsf{e})(p)) \in cl(\varphi)$, see Definition 19 on 22. Similarly, we expect the postconditions to turn up when measuring the complexity of formulas. The inductive proof of the truth lemma for the logic is by induction on the complexity of formulas (in the closure set). It is derivable that $c([\mathsf{U},\mathsf{e}]p) > c(\mathsf{pre}(\mathsf{e}) \rightarrow \mathsf{post}_{\mathsf{e}}(p))$, see Lemma 26, item 2, on page 26. These are the only parts of the completeness proof where postconditions occur.

4 Semantic results

The following subsections present some semantic peculiarities of the logic. They may be helpful in relating this approach to other and existing approaches combining ontic and epistemic change, and may further be helpful (different 'normal forms' are defined) towards future tool development.

4.1 Arbitrary belief change

Let (M,s) and (M',s') be arbitrary finite epistemic states for the same set of atoms and agents, with $M = (S,R,V)$ and $M' = (S',R',V')$. Surprisingly enough, there is 'almost always' an update transforming the former into the latter. 'Almost always': there are two constraints, both of a somewhat technical nature. Firstly, for the set of agents with non-empty access in M' there must be a submodel of M containing actual state s that is serial for those agents. In other words (slightly less precise), if an agent initially believes everything ('is crazy') you cannot change his beliefs, but otherwise you can. This seems reasonable. Secondly, M and M' should only differ in the value of a finite number of atoms (called the *relevant* atoms for this purpose). This is required, because we can only change a finite number of atoms in the postconditions. This also seems reasonable: as both models are finite, the agents can only be uncertain about the value of a *finite* number of atoms (in the combined models M and M'); in other words, they are 'not interested' in the value of the remaining atoms.

For presentation purposes we initially assume that all agents consider the actual state s of M a possibility (as in all $S5$ models, such as Kripke models representing interpreted systems), thus satisfying the first of the two constraints above: the serial submodel required is then the singleton model consisting of s, accessible to all agents. The update transforming (M, s) into (M', s') can be seen as the composition of two intuitively more appealing updates. That will make clear how we can also describe the required update in one stroke.

In the first step we get rid of the structure of M. As the epistemic state (M, s) is finite, it has a characteristic formula $\delta_{(M,s)}$ [6, 20]. We let the agents publicly learn that characteristic formula. This event is represented by the singleton update (U, e) defined as

$$((\{\mathsf{e}\}, \mathsf{R}, \mathsf{pre}, \mathsf{post}), \mathsf{e}) \text{ with } \mathsf{pre}(\mathsf{e}) = \delta_{(M,s)}, \text{ for all } a : (\mathsf{e}, \mathsf{e}) \in \mathsf{R}(a), \text{ and } \mathsf{post}(\mathsf{e}) = \emptyset$$

In other words, the structure of the current epistemic state is being publicly announced. The resulting epistemic state is, of course, also singleton, or bisimilar to a singleton epistemic state, as $\delta_{(M,s)}$ holds in all states in M bisimilar to s. (For the definition of bisimulation, see [7].) Without loss of generality assume that it is singleton. Its domain is $\{(s, \mathsf{e})\}$. This pair (s, e) is accessible to itself because for all agents, $(s, s) \in R(a)$ (all agents consider the actual state s a possibility), and $(\mathsf{e}, \mathsf{e}) \in \mathsf{R}(a)$. The valuation of propositional variables in this intermediate state are those of state s in M. This does not matter: we will not use that valuation.

Now proceed to the second step. In the epistemic state wherein the agents have common knowledge of the facts in s, the agents learn the structure of the resulting epistemic state M' and their part in it by executing update (U', s') defined as

$$((S', R', \mathsf{pre}', \mathsf{post}'), s') \text{ with for } t' \in S' : \mathsf{pre}'(t') = \top \text{ and for relevant } p : \mathsf{post}'(t')(p) = \top \text{ iff } t' \in V'(p)$$

where the domain S' and the accessibility relation R' of U' are precisely those of M', the resulting final epistemic model. The postcondition post' is well-defined, as only a finite number of atoms (the relevant atoms) is considered. Because we execute this update in a singleton model with public access, and because it is executable for every event t', the resulting epistemic state has the same structure as the update: it returns S' and R' again. The postcondition delivers the required valuation of atoms in the final model: for each *event* t' in U' and for all revelant atoms p, p become true in t' if p is true in *state* t' in M' ($\mathsf{post}(t')(p) = \top$), else p becomes false. The value of irrelevant atoms remains the same.

We combine these two updates into one by requiring the precondition of the first and the postcondition of the second. Consider U'_r that is exactly as U' except that in all events t' in its domain the precondition is not $\top$ but $\delta_{(M,s)}$: the characteristic formula of *the point* s of (M, s). Update (U'_r, s') does the job: epistemic state $(M \otimes \mathsf{U}'_r, (s, s'))$ is isomorphic to (M', s'). This will be Corollary 11 of our more general result, to follow.

Now consider the more general case that the set of all agents with non-empty access in M', henceforth called B, are serial in a submodel of M that contains s, henceforth called M^{ser}, with $M^{\mathsf{ser}} = (S^{\mathsf{ser}}, R^{\mathsf{ser}}, V^{\mathsf{ser}})$. In other words: at least all agents who finally have consistent beliefs in some states, initially have consistent beliefs in all states. The construction above will no longer work: if the actual state is not considered possible by an agent, then that agent will have empty access in $(M \otimes \mathsf{U}'_r)$ but not in M'. But if we relax the precondition $\delta_{(M,s)}$, for the point s of (M, s), to the disjunction $\bigvee_{u \in S^{\mathsf{ser}}} \delta_{(M,u)}$ that will now carry along the serial subdomain S^{ser}, the construction will work because an agent can then always imagine *some* state wherein the update has been executed, even it that is not the actual state. Formally, we get the update $(\mathsf{U}'', (s, s')) = ((\mathsf{E}'', \mathsf{R}'', \mathsf{pre}'', \mathsf{post}''), (s, s'))$ such that (for arbitrary agents a and arbitrary relevant atoms p):

$$\begin{array}{lcl} \mathsf{E}'' & = & S' \\ (t', u') \in \mathsf{R}''(a) & \text{iff} & (t', u') \in R'(a) \\ \mathsf{pre}''(t') & = & \bigvee_{u \in S^{\mathsf{ser}}} \delta_{(M,u)} \\ \mathsf{post}''(t')(p) & = & \begin{cases} \top & \text{if } t' \in V'(p) \\ \bot & \text{otherwise} \end{cases} \end{array}$$

The epistemic state $(M \otimes \mathsf{U}'', (s, s'))$ is bisimilar to the stipulated (M', s'), which is the desired result. It will typically not be isomorphic: $M \otimes \mathsf{U}''$ can be seen as consisting of a number of copies of M' (namely $|S^{\mathsf{ser}}|$ copies) 'with the accessibility relations just right to establish the bisimulation'. One copy will be typically 'not enough', namely when the state tail in M to which that copy corresponds, lacks access for some agents. This access will then also be 'missing' between the states of $(\{\mathsf{tail}\} \times S')$. But because of seriality one of the other M' copies will now make up for this lack: there is a $u \in S^{\mathsf{ser}}$ such that $(t, u) \in R(a)$, which will establish access when required (as in the proof of the following proposition).

Proposition 10 Given (M, s), (M', s'), and U'' as above. Then $\mathfrak{R} : ((M \otimes \mathsf{U}''), (s, s')) \underline{\leftrightarrow} (M', s')$ by way of, for all $t \in S^{\mathsf{ser}}$ and $t' \in S'$: $\mathfrak{R}(t, t') = t'$. ⊣

Proof Let $R^{\otimes}$ be the accessibility relation and $V^{\otimes}$ the valuation in $(M \otimes \mathsf{U}'')$.

atoms: For an arbitrary relevant atom p: $(t, t') \in V^{\otimes}(p)$ iff $(M, t) \models \mathsf{post}''(t')(p)$, and by definition of post'' we also have that $(M, t) \models \mathsf{post}''(t')(p)$ iff $t' \in V'(p)$. Irrelevant atoms do not change value.

forth: Let $((t_1, t'_1), (t_2, t'_2)) \in R^{\otimes}(a)$ and $((t_1, t'_1), t'_1) \in \mathfrak{R}$. From $((t_1, t'_1), (t_2, t'_2)) \in R^{\otimes}(a)$ follows $(t'_1, t'_2) \in R'(a)$. By definition of $\mathfrak{R}$ we also have $((t_2, t'_2), t'_2) \in \mathfrak{R}$.

back: Let $((t_1, t'_1), t'_1) \in \mathfrak{R}$ and $(t'_1, t'_2) \in R'(a)$. As M^{ser} is serial for a, and $t_1 \in S^{\mathsf{ser}}$, there must be a t_2 such that $(t_1, t_2) \in R(a)$. As $(M, t_2) \models \bigvee_{t \in \mathrm{dom}(M^{\mathsf{ser}})} \delta_{(M,t)}$ (because t_2 is one of those t) we have that $(t_2, t'_2) \in \mathrm{dom}(M \otimes \mathsf{U}'')$. From that, $(t_1, t_2) \in R(a)$, and $(t'_1, t'_2) \in R'(a)$, follows that $((t_1, t'_1), (t_2, t'_2)) \in R^{\otimes}(a)$. By definition of $\mathfrak{R}$ we also have $((t_2, t'_2), t'_2) \in \mathfrak{R}$.

Note that we keep the states outside the serial submodel M^{ser} out of the bisimulation. Without the seriality constraint the 'back' condition of the bisimilarity cannot be shown: given a $((t_1, t'_1), t'_1) \in \mathfrak{R}$ and $(t'_1, t'_2) \in R'(a)$, but where t_1 has no outgoing arrow for a, the required a-accessible pair from (t_1, t'_1) does not exist.

Corollary 11 Given (M, s), (M', s'), and U'_r as above. Assume that M is a bisimulation contraction. Then $(M \otimes \mathsf{U}'') \cong M'$. ⊣

Proof In this special case we have that $(t, t') \in \mathrm{dom}(M \otimes \mathsf{U}'_r)$ iff $(M, t) \models \mathsf{pre}'(t')$ iff $(M, t) \models \delta_{(M,s)}$ for the point s of (M, s). As the last is only the case when $t = s$ (as M is a bisimulation contraction), we end up with a domain consisting of all pairs (s, t') for all $t' \in S'$, a 1-1-correspondence. The bisimulation $\mathfrak{R}$ above becomes the isomorphism $\mathfrak{I}(s, t') = t'$.

Corollary 12 For arbitrary finite epistemic states (M, s) and (M', s') there is an update (U, e) transforming the first into the second (as long the two constraints are satisfied). ⊣

Corollary 13 Given a finite epistemic state (M, s), and a satisfiable formula φ. (As long as M has a serial submodel for all agents occurring in φ) There is an update *realizing* φ, i.e., there is a (U, e) such that $(M, s) \models \langle \mathsf{U}, \mathsf{e} \rangle \varphi$. ⊣

Using completeness of the logic, this further implies that all consistent formulas can be realized in any given finite model with distributed knowledge. We find this result both weak and strong: it is strong because any conceivable (i.e., using the same propositional letters and set of agents) formal specification can be made true whatever the initial information state. At the same time, it is weak: the current information state does apparently not give any constraints on future developments of the system, or, in the opposite direction, any clue on the sequence of events resulting in it; the ability to change the value of atomic propositions arbitrarily gives too much freedom. Of course, if one restricts the events to certain types, the amount of change is constrained. The example scenarios in Section 5 illustrate that.

4.2 Postconditions true and false only

The postconditions for propositional atoms can be entirely simulated by the postconditions true or false for propositional atoms. For a simple example, the public assignment $p := \varphi$ can be simulated by a two-point update $\underline{\mathsf{e}}$——A——f (i.e., all agents in A cannot distinguish e from f) such that $\mathsf{pre}(\mathsf{e}) = \varphi$, $\mathsf{post}(\mathsf{e})(p) = \top$, $\mathsf{pre}(\mathsf{f}) = \neg\varphi$, $\mathsf{post}(\mathsf{f})(p) = \bot$. In the public assignment $(p := \varphi, q := \psi)$ to two atoms p and q we would need a *four*-point update to simulate it, to distinguish all *four* ways to combine the values of two independent atoms.

The general construction consists of doing likewise in every event e of an update model. For each e we make as many copies as the cardinality of the powerset of the range of the postcondition associated with that event. Below, the set $\{0,1\}^{\mathrm{dom}(\mathsf{post}(\mathsf{e}))}$ represents that powerset.

Definition 14 (Update model $\mathsf{U}^{\top\bot}$) Given is update model $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$. Then $\mathsf{U}^{\top\bot} = (\mathsf{E}^{\top\bot}, \mathsf{R}^{\top\bot}, \mathsf{pre}^{\top\bot}, \mathsf{post}^{\top\bot})$ is a *normal update model* with

- $\mathsf{E}^{\top\bot} = \bigcup_{\mathsf{e}\in\mathsf{E}}\{(\mathsf{e}, f) \mid f \in \{0,1\}^{\mathrm{dom}(\mathsf{post}(\mathsf{e}))}\}$
- $((\mathsf{e}, f), (\mathsf{e}', f')) \in \mathsf{R}^{\top\bot}(a)$ iff $(\mathsf{e}, \mathsf{e}') \in \mathsf{R}(a)$
- $\mathsf{pre}^{\top\bot}(\mathsf{e}, f) = \mathsf{pre}(\mathsf{e}) \wedge \bigwedge_{f(p)=1} \mathsf{post}(\mathsf{e})(p) \wedge \bigwedge_{f(p)=0} \neg\mathsf{post}(\mathsf{e})(p)$
- $\mathsf{post}^{\top\bot}(\mathsf{e}, f)(p) = \begin{cases} \top & \text{if } f(p) = 1 \\ \bot & \text{if } f(p) = 0 \end{cases}$ ⊣

Proposition 15 Given an epistemic model $M = (S, R, V)$ and an update model $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$ with normal update model $\mathsf{U}^{\top\bot}$ defined as above. Then $(M \otimes \mathsf{U}) \underline{\leftrightarrow} (M \otimes \mathsf{U}^{\top\bot})$. ⊣

Proof We show that the relation $\mathfrak{R} : (M \otimes \mathsf{U}) \underline{\leftrightarrow} (M \otimes \mathsf{U}^{\top\bot})$ defined as

$$((s, \mathsf{e}), (s, \mathsf{e}, f)) \in \mathfrak{R} \text{ iff } (M, s) \models \mathsf{pre}^{\top\bot}(\mathsf{e}, f)$$

is a bisimulation. Below, the accessibility relations in $(M \otimes \mathsf{U})$ and $(M \otimes \mathsf{U}^{\top\bot})$ are also written as $R(a)$.

atoms

We have to show that for all atoms p, $(M, s) \models \mathsf{post}(\mathsf{e})(p) \leftrightarrow \mathsf{post}^{\top\bot}(\mathsf{e}, f)(p)$. From the definition of $\mathsf{post}^{\top\bot}$ it follows that

$$\mathsf{post}^{\top\bot}(\mathsf{e}, f)(p) \text{ iff } f(p) = 1 .$$

From $(M, s) \models \mathsf{pre}^{\top\bot}(\mathsf{e}, f)$ and the definition of $\mathsf{pre}^{\top\bot}$ follows that

$$(M, s) \models \mathsf{post}(\mathsf{e})(p) \text{ iff } f(p) = 1 .$$

Therefore

$$(M, s) \models \mathsf{post}(\mathsf{e})(p) \leftrightarrow \mathsf{post}^{\top\bot}(\mathsf{e}, f)(p) .$$

forth

Assume that $((s, \mathsf{e}), (s', \mathsf{e}')) \in R(a)$ and that $((s, \mathsf{e}), (s, \mathsf{e}, f)) \in \mathfrak{R}$. Let $f' : \mathrm{dom}(\mathsf{post}(\mathsf{e}')) \to \{0,1\}$ be the function such that

$$f'(p) = \begin{cases} 1 & \text{if } \mathsf{post}(\mathsf{e}')(p) = \varphi \text{ and } (M, s') \models \varphi \\ 0 & \text{if } \mathsf{post}(\mathsf{e}')(p) = \varphi \text{ and } (M, s') \not\models \varphi \end{cases}$$

Therefore $(M, s') \models \mathsf{pre}^{\top\bot}(\mathsf{e}', f')$. Therefore $((s', \mathsf{e}'), (s', \mathsf{e}', f')) \in \mathfrak{R}$. From $((s, \mathsf{e}), (s', \mathsf{e}')) \in R^{\top\bot}(a)$ follows $(s, s') \in R(a)$ and $(\mathsf{e}, \mathsf{e}') \in \mathsf{R}(a)$. From $(\mathsf{e}, \mathsf{e}') \in \mathsf{R}(a)$ and the definition of access $R^{\top\bot}$ follows $((\mathsf{e}, f), (\mathsf{e}', f')) \in R^{\top\bot}(a)$. From $(s, s') \in R(a)$ and $((\mathsf{e}, f), (\mathsf{e}', f')) \in R^{\top\bot}(a)$ follows $((s, \mathsf{e}, f), (s', \mathsf{e}', f')) \in R(a)$.

back

Suppose $((s, \mathsf{e}), (s, \mathsf{e}, f)) \in \mathfrak{R}$ and $(s, \mathsf{e}, f), (s', \mathsf{e}', f') \in R(a)$. From the last follows $(s, s') \in R(a)$ and $((\mathsf{e}, f), (\mathsf{e}', f')) \in \mathsf{R}^{\top\bot}(a)$, therefore also $(\mathsf{e}, \mathsf{e}') \in \mathsf{R}(a)$. Therefore $((s, \mathsf{e}), (s, \mathsf{e}')) \in R(a)$. That $((s', \mathsf{e}'), (s', \mathsf{e}', f')) \in \mathfrak{R}$, is established as in **forth**.

Corollary 16 The logic of change with postconditions true and false only is just as expressive as the logic of change with arbitrary postconditions. ⊣

Of course the reader realizes, that although it is possible to use postconditions true and false only, this is highly unpractical in modelling actual situations: the descriptions of updates become cumbersome and lengthy, and lack intuitive appeal.

A similar transformation result can *not* be established for *public* (epistemic or ontic) change (as in [13]). If public assignments could only be to true and to false, then updates with assignments always result in models wherein the assigned atoms are true *throughout* the model, or false *throughout* the model. For example, there is no transformation of $\underline{p}$——$\neg p$ into p——$\underline{\neg p}$ with *public* assignments and announcements only.

A transformation result as in Proposition 15 immediately gives an expressivity result as in Corollary 16 for the languages concerned. It is also tempting to see such a transformation result as a different kind of expressivity result. In two-sorted languages such as the one we consider in this paper one can then distinguish between the expressivity of two kinds of syntactic objects. A formula (φ) corresponds to a class of models that satisfy that formula, and a modality (α) corresponds to a relation on the class of models. The ability to express more relations does not necessarily lead to the ability to express more classes of models, nor vice versa. For example, what formulas are concerned, epistemic logic without common knowledge is equally expressive as public announcement logic without common knowledge, and even as public announcement and assignment logic without common knowledge [13]. But, as we have now seen, more and more relations between the models can be expressed.

4.3 Single assignments only

Consider update model e——a——f for a single agent a and for two atoms p_1 and p_2 such that in e, if φ_1 then $p_1 := \varphi_2$ and $p_2 := \varphi_3$, and in f, if φ_4 then $p_1 := \varphi_5$ and $p_2 := \varphi_6$. Can we also do the assignments one by one? In other words, does this update correspond to a sequence of updates consisting of events g in which at most one atom is assigned a value: the cardinality of $\text{dom}(\mathsf{g})$ is at most 1. This is possible! *First* we 'store' the value (in a given model (M, s) wherein this update is executed) of all preconditions and postconditions in fresh atomic variables, by public assignments. This can be in arbitrary order, so we do it in the order of the φ_i. This is the sequence of six public assignments $q_1 := \varphi_1$, $q_2 := \varphi_2$, $q_3 := \varphi_3$, $q_4 := \varphi_4$, $q_5 := \varphi_5$, and $q_6 := \varphi_6$. Note that such public assignments do not change the structure of the underlying model. *Next* we execute the original update but without postconditions. This is e'——a——f' with $\mathsf{pre}(\mathsf{e}') = \mathsf{pre}(\mathsf{e}) = \varphi_1$ and $\mathsf{pre}(\mathsf{f}') = \mathsf{pre}(\mathsf{f}) = \varphi_4$ and with $\mathsf{post}(\mathsf{e}') = \mathsf{post}(\mathsf{f}') = \emptyset$. Note that q_1 remains true whenever e' was executed, because q_1 was set to be true whenever φ_1 was true, the precondition of both e and e'. Similarly, q_4 remains true whenever f' was executed. We have now arrived at the final structure of the model, just not at the proper valuations of atoms.

Finally, the postconditions are set to their required value, *conditional* to the execution of the event with which they are associated. Agent a must not be aware of those conditions (the agent cannot distinguish between e' and f'). Therefore we cannot model this as a public action. The way out of our predicament is a number of two-event update models, namely one for each postcondition of each event in the original update. One of these two events has as its precondition the fresh atom associated with an event in the original update, and the other event its negation, and agent a cannot distinguish between both. The four required updates are

- e_1——a——e_1' with in e_1, if q_1 then $p_2 := q_2$ and in e_1', if $\neg q_1$ then $\emptyset$
- e_2——a——e_2' with in e_2, if q_1 then $p_3 := q_3$ and in e_2', if $\neg q_1$ then $\emptyset$

- e_3——a——e_3' with in e_3, if q_4 then $p_5 := q_5$ and in e_3', if $\neg q_4$ then $\emptyset$
- e_4——a——e_4' with in e_4, if q_4 then $p_6 := q_6$ and in e_4', if $\neg q_4$ then $\emptyset$

Now, we are done. These four final updates do not change the structure of the model, when executed. Therefore, now having set the postconditions right, the composition of all these constructs is *isomorphic* to the original update model! The general construction will hopefully be sufficiently clear from this simple example.

Definition 17 (Update model $\mathsf{U}^{\mathsf{one}}$) Given an update model $\mathsf{U} = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post})$, update model $\mathsf{U}^{\mathsf{one}}$ is the composition of the following update models: *First* perform $\Sigma_{\mathsf{e}\in\mathsf{E}}|\text{dom}(\mathsf{post}(\mathsf{e})) + 1|$ public assignments for fresh variables $q, \ldots$, namely for each $\mathsf{e} \in \mathsf{E}$, $q_0^{\mathsf{e}} := \mathsf{pre}(\mathsf{e})$, and for all $p_1, \ldots, p_n \in \text{dom}(\mathsf{post}(\mathsf{e}))$, $q_1^{\mathsf{e}} := \mathsf{post}(\mathsf{e})(p_1)$, ..., $q_n^{\mathsf{e}} := \mathsf{post}(\mathsf{e})(p_n)$. *Then* execute U but with empty postconditions, i.e., execute $\mathsf{U}' = (\mathsf{E}, \mathsf{R}, \mathsf{pre}, \mathsf{post}')$ with $\mathsf{post}'(\mathsf{e}) = \emptyset$ for all $\mathsf{e} \in \mathsf{E}$. *Finally*, execute $\Sigma_{\mathsf{e}\in\mathsf{E}}|\text{dom}(\mathsf{post}(\mathsf{e}))|$ two-event update models with universal access for all agents wherein for each event just one of its postconditions is set to its required value, by way of the auxiliary atoms. For example, for $\mathsf{e} \in \mathsf{E}$ as above the first executed update is e_1——A——e_2 with in e_1, if q_0^{e}, then $p_1 := q_1^{\mathsf{e}}$, and in e_2, if $\neg q_0^{\mathsf{e}}$ then $\emptyset$. ⊣

Without proof the following proposition will be clear:

Proposition 18 Given epistemic model M and update model U executable in M. Then $\mathsf{U}^{\mathsf{one}}$ is isomorphic to U, and $(M \otimes \mathsf{U}^{\mathsf{one}})$ is isomorphic to $(M \otimes \mathsf{U})$. ⊣

This result brings the logic closer to the proposals in [4, 26] wherein only one atom is simultaneously assigned a value.

5 Applications

In this section we apply the logic to model multi-agent system dynamics in two fairly general settings. Subsection 5.1 models various game actions for card players, such as showing, drawing, and swapping cards. Such precise dynamics are essential prerequisites to the further analysis of such games, e.g., computing equilibrium strategies. Subsection 5.2 models a protocol that solves the well-known 'one hundred prisoners and a lightbulb' riddle. The events involved are that a prisoner may turn that light on, or off, or make an announcement, depending on the state of the light. This serves as an example of the applicability of this logic to modelling other protocols with complex communicative features.

5.1 Card game actions

Consider a deck of two Wheat, two Flax, and two Rye cards (w, x, y). Wheat, Flax and Rye are called the *commodities*. Three players Anne, Bill, and Cath (a, b, and c) each draw two cards from the stack. Initially, given a deal of cards, it is common knowledge what the deck of cards is, that all players hold two cards, and that all players (only) know their own cards. For the card deal where Anne holds a Wheat and a Flax card, Bill a Wheat and a Rye card, and Cath a Flax and a Rye card, we write $wx.wy.xy$, and so on. As the cards in one's hand are unordered, $wx.wy.xy$ is the same deal of cards as $xw.wy.xy$, but for improved readability we will always list cards in a hand in alphabetical order. There are certain game actions that result in players exchanging cards. This is called *trading* of the corresponding commodities. Players attempt to get two cards of the same suit. That is called establishing a *corner* in a commodity. Subject to game rules that are non-essential for our exposition, the first player to declare a corner in any commodity, wins the game. For example, given deal $wx.wy.xy$, after Anne swaps her Wheat card for Bill's Rye card, Bill achieves a corner in Wheat, and wins. Of course, players can already achieve a corner when the cards are dealt. This six-card scenario is a simplification of the 'Pit' card game that simulates the trading pit of a stock exchange [15, 23, 25]; the full game consists of 74 cards: 8 commodities of each 9 cards, and two special cards.

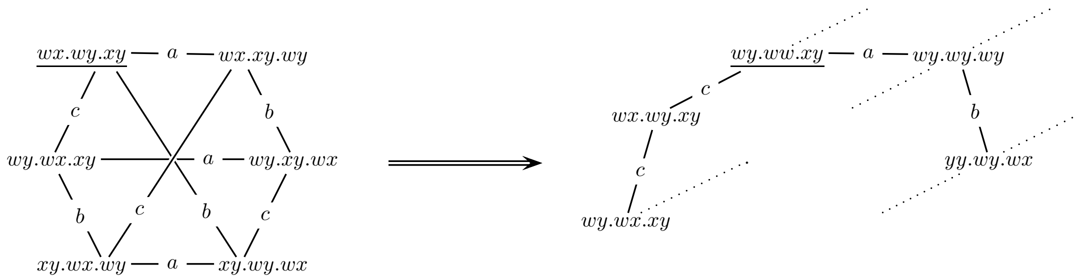


Figure 2: On the left is the game state after the cards have been dealt and Anne received Wheat and Flax, Bill received Wheat and Rye, and Cath received Flax and Rye. On the right is part of the game state that results if Anne trades her Wheat card for Bill's Rye card: only states resulting from trading Wheat for Wheat, and (what really happened) Wheat for Rye, are present. The actual deal of cards is underlined. In the figures, assume reflexivity and transitivity of access. The dotted lines suggest that some states are indistinguishable for Cath from yet other states but not present in the picture.

An initial game state wherein players only know their own cards and nobody has won the game yet, can be modelled as an epistemic state. There are six such card deals. Assume that the actual deal is $wx.wy.xy$. The (hexagonally shaped) epistemic state $(Pit, wx.wy.xy)$ in Figure 2 pictures what players know about each other. All six card deals have to occur to describe the epistemic information present, e.g., Anne cannot distinguish actual deal $wx.wy.xy$ from deal $wx.xy.wy$, the other deal wherein Anne holds Wheat and Flax. But if the deal had been $wx.xy.wy$, Bill could not have distinguished *that* deal from $wy.xy.wx$, wherein Anne holds Wheat and Rye. Therefore, Anne considers it possible that Bill considers it possible that she holds Rye, even though this is actually not the case.

The event wherein Anne and Bill swap one of their cards involves both epistemic and ontic change. Assume that, given deal $wx.wy.xy$,

> Anne swaps her Wheat card for Bill's Rye card.

This informal description is not specific enough to be modelled as an update.

In the first place, the role of Cath in the event is not specified. The event where Cath is unaware of the swap taking place, is different from the event where Cath observes the swap and where all agents are commonly aware of this. If Cath had been unaware of the event, she would be mistaken about the actual state of the world. For example, she would incorrectly still believe that neither Anne nor Bill has a corner in a commodity, whereas Bill holds two Wheat cards after the trade (and we assume that he has not yet so declared). It is hard to conceive of such a scenario as a *game*: even in imperfect information games, such as Pit, a basic condition of *fairness* must be fulfilled for players to be able to act rationally. This means that all events should at least be partially observable, and that 'what actually happened' should be considered a possibility for all players. We therefore assume, for now, that Cath learns that Anne and Bill swap one of their cards, but not which card it is. (The 'private swap' will be modelled later, as another example.)

Anne and Bill's roles in the event are also underspecified. Anne may knowingly *choose* a card to hand to Bill (the obvious interpretation), or blindly *draw* one of her cards to hand to Bill. The latter is not obvious, given this description, but becomes more so if we see it as Bill drawing (therefore blindly) one of Anne's cards. For now, assume the obvious. Another specification issue is that we may think of Bill as receiving Anne's Wheat card facedown and only then, in a subsequent action, picking it up. From our modelling perspective, Bill already can be said to *own* the card after he has been handed it, but before he has picked it up he does not yet *know* that he owns it. We first assume that players immediately 'see' the card they are being traded (in this case maybe not the most obvious choice, but the simplest one to model). In other words: Anne and Bill *jointly* learn the new ownership of both cards.

To describe this multi-agent system and its dynamics, assume a propositional language for three agents a, b, c and with atoms u_a^n expressing that Anne holds n cards of suit u. For example, w_a^2 expresses that Anne holds two Wheat cards. In the event where Anne and Bill swap Wheat for Rye, Anne gets one *less* Wheat card, Bill gets one *more* Wheat card, Bill gets one *less* Rye card, and Anne gets one *more* Rye card. In the update model we have to distinguish separate events for each card deal wherein this swap can take place, i.e., corresponding to $wx.wy.xy$, $wx.xy.wy$, and $wy.xy.wx$ (in general this depends on a feature of the local states of the card swapping agents only, namely for both agents on the number of Wheat and Rye cards in their hands, in this specific case that information is sufficient to determine the entire card deal). In case the card deal is $wx.wy.xy$ the precondition and postcondition are

If $(w_a^1 \wedge y_a^0 \wedge w_b^1 \wedge y_b^1)$, then $w_a^1 := \bot$ and $w_a^0 := \top$ and $w_b^1 := \bot$ and $w_b^2 := \top$ and $y_a^0 := \bot$ and $y_a^1 := \top$ and $y_b^1 := \bot$ and $y_b^0 := \top$.

We name this event $swap_{ab}^{wx.wy.xy}(w, y)$. If two cards of the same suit are swapped, a simpler description is sufficient. For example, the event wherein Anne and Bill swap Wheat given deal $wx.wy.xy$ is described as $swap_{ab}^{wx.wy.xy}(w, w)$ with (the same) precondition and (empty) postcondition

If $(w_a^1 \wedge y_a^0 \wedge w_b^1 \wedge y_b^1)$, then $\emptyset$

From the point of view of an actual card deal, there are always four different ways to exchange a single card, namely for each agent either the one or the other card. All of these are clearly different for Anne and Bill, because they either give or receive a different card (we assumed that they know which card they give and see which card they receive). None of these are different for Cath. For different card deals, card swapping events are indistinguishable if those deals were indistinguishable. For example, the event where (Anne and Bill swap Wheat and Rye given $wx.wy.xy$) is indistinguishable for Anne from the event where (Anne and Bill swap Wheat and Rye given $wx.xy.wy$), because card deals $wx.wy.xy$ and $wx.xy.wy$ are the same for Anne.

Therefore, the update model for Anne swapping her Wheat card for Bill's Rye card consists of 24 events. The preconditions and postconditions of the events are as above. The accessibility relations are defined as, for deals $d, d' \in \text{dom}(Pit) = \{wx.wy.xy, wx.xy.wy, \dots\}$ and cards $q, q', q_1, q_1' \in \{w, x, y\}$, and accessibility relations $R(a), R(b), R(c)$ in the epistemic model Pit:

$$\begin{array}{lll} (swap_{ab}^{d}(q,q'), swap_{ab}^{d'}(q_1,q_1')) \in \mathsf{R}(a) & \text{iff} & (d,d') \in R(a) \text{ and } q = q_1 \text{ and } q' = q_1' \\ (swap_{ab}^{d}(q,q'), swap_{ab}^{d'}(q_1,q_1')) \in \mathsf{R}(b) & \text{iff} & (d,d') \in R(b) \text{ and } q = q_1 \text{ and } q' = q_1' \\ (swap_{ab}^{d}(q,q'), swap_{ab}^{d'}(q_1,q_1')) \in \mathsf{R}(c) & \text{iff} & (d,d') \in R(c) \end{array}$$

We name the update model *Swap*. The event of Anne and Bill swapping Wheat for Rye has therefore been modelled as update $(Swap, swap_{ab}^{d}(w, y))$. The result of executing this update model in epistemic state $(Pit, wx.wy.xy)$ has the same structure as the update model, as the preconditions are unique for a given state, and as access between events in the update model copies that in the epistemic state. It has been partially visualized in Figure 2. An intuitive way to see the update and the resulting structure in Figure 2, is as a restricted product of the Pit model and *nine* card swapping events $swap(q, q_1)$, namely for each combination of the three different cards. Figure 2 then shows just two of those copies, namely for $swap(w, y)$ and $swap(w, w)$. For example, the event $swap(w, y)$ 'stands for' the three events $swap_{ab}^{wx.wy.xy}(w, y)$, $swap_{ab}^{wy.wx.xy}(w, y)$, and $swap_{ab}^{wx.wy.xy}(w, y)$.

Why did we not define such $swap(q, q_1)$ as updates in their own right, in the first place? Although intuitive, this is not suppported by our modelling language. We would like to say that the postconditions are 'Anne gets *one less* Wheat card, and Bill gets *one more* Wheat card,' and similarly for Rye. But instead, we only *can* demand that in case Bill already had a Wheat card (extra precondition), *then* he now has two, etc. Incidentally, we can also *either* add non-deterministic choice to the update language by notational abbreviation, as $[\alpha \cup \beta]\varphi \leftrightarrow ([\alpha]\varphi \wedge [\beta]\varphi)$ (this corresponds to taking the *union* of the epistemic state transformations induced by α and β), *or* allow multi-pointed update models. We can then define, in the update language, $swap(w, y) = swap_{ab}^{wx.wy.xy}(w, y) \cup swap_{ab}^{wy.wx.xy}(w, y) \cup swap_{ab}^{wx.wy.xy}(w, y)$, or, respectively, see $swap(w, y)$ as a three-pointed $Swap$ model.

The case where Anne does not choose her card but Bill blindly draws one of Anne's can also be modelled as an update. The accessibility for Anne then expresses that she is unaware which of her cards has been drawn:

$$(swap_{ab}^{d}(q, q'), swap_{ab}^{d'}(q_1, q_1')) \in \mathsf{R}(a) \quad \text{iff} \quad (d, d') \in R(a) \text{ and } q' = q_1'$$

This is somewhat counterintuitive when we still suppose that Anne observes which card she receives from Bill. (We'd have to imagine Bill blindly drawing one of Anne's cards, Anne putting her remaining card facedown on the table, and receiving the card Bill gives her faceup.) A more realistic setting is then that Bill draws one of Anne's card and 'pushes the card he gives to Anne facedown towards her'. At that stage Anne can already be said to own the card, but not yet to know that. All four swapping actions for a given deal are indistinguishable for Anne (as they were and still are for Cath).

Yet another event is where Anne chooses a card to show to Bill, and receives Bill's card facedown (before she picks it up). Access is now

$$(swap_{ab}^{d}(q, q'), swap_{ab}^{d'}(q_1, q_1')) \in \mathsf{R}(a) \quad \text{iff} \quad (d, d') \in R(a) \text{ and } q = q_1$$

Obviously, all these variables can be applied to Bill as well.

Picking up a card The action of picking up a card after it has been handed to you, has another description (see, using another dialect of the language, [25]). One aspect is interesting to observe in the current context. Imagine that given deal $wx.wy.xy$ after Anne and Bill swapping Wheat and Rye, Bill receives the card facedown in the following way: after having laid down his remaining card (a Wheat card) facedown on the table, Anne puts her Wheat card facedown on top of it or under it in a way that Bill cannot see that. He now picks up his two cards. How does Bill know which of the two Wheat cards that he then holds, is the received card? He does not know, but he does not care either. By looking at his cards after the swap, he effectively *learns* that he holds two Wheat cards (which was already true after having received the card), and after that event he then *knows* that he holds two Wheat cards. A neat way to express that he learns the suit of the card he received, is to say that there is a suit for which he learns to have one more card than he knows. This makes sense in the general Pit game setting wherein one only is allowed to trade a certain number of cards of the same suit. This is formalized in our setting by an update (U, e) with an event with precondition $\mathsf{pre}(\mathsf{e}) = w_b^2 \wedge \neg[b]w_b^2$ (and empty postcondition), as a result of which Bill then knows to hold two Wheat cards, i.e., $[\mathsf{U}, \mathsf{e}][b]w_b^2$.

Private swap The event where Anne and Bill swap Wheat for Rye but where Cath is unaware of the event is modelled by a four-event update with events $swap_{ab}^{wx.wy.xy}(w, y)$, $swap_{ab}^{wy.wx.xy}(w, y)$, $swap_{ab}^{wx.wy.xy}(w, y)$ and skip, such that for Anne and Bill access among the *swap* events is as already discussed (including all variations), but where 'Cath thinks nothing happens': for the deals d in the three swap events: $(swap_{ab}^{d}(w, y), \mathsf{skip}) \in R(c)$, and $(\mathsf{skip}, \mathsf{skip}) \in R(a), R(b), R(c)$.

The precise description of card game dynamics is a prerequisite to compute optimal strategies to play such games [23, 8]. Card deals are also frequently used as standard representation for cryptographic bit exchange protocols [9, 18], where communications / transmissions should from our current perspective be seen as public announcements and/or assignments. Our more fine-grained analysis of events may contribute to the description and verification of more complex protocols that also include non-public events. The next subsection treats such a protocol, although not in a setting of card deals.

5.2 One hundred prisoners and a lightbulb

The following riddle appears to have been around for at least five years[4].

[4]We made some investigation on the puzzle's origin. Our source was the ESSLLI 2003 logic summerschool in Vienna, where it was—apparently incorrectly—said to originate with Moshe Vardi. We did not find informal

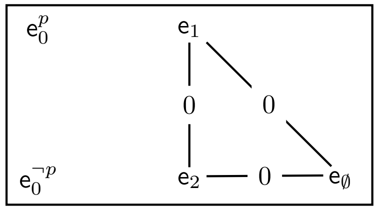


Figure 3: The update model for an interrogation for the situation with three prisoners. See accompanying text for explanations.

> A group of 100 prisoners, all together in the prison dining area, are told that they will be all put in isolation cells and then will be interrogated one by one in a room containing a light with an on/off switch. The prisoners may communicate with one another by toggling the light-switch (and that the only way in which they can communicate). The light is initially switched off. There is no fixed order of interrogation, or interval between interrogations, and the same prisoner may be interrogated again at any stage. When interrogated, a prisoner can either do nothing, or toggle the light-switch, or announce that all prisoners have been interrogated. If that announcement is true, the prisoners will (all) be set free, but if it is false, they will all be executed. While still in the dining room, and before the prisoners go to their isolation cells (forever), can the prisoners agree on a protocol that will set them free?

Of course, the answer is: "Yes, they can." We solve the riddle for an arbitrary number of n prisoners. For $n = 1$ and $n = 2$ it is trivial. For $n > 2$ a protocol is as follows:

> The n prisoners appoint one amongst them as the 'counter'. All non-counting prisoners follow the following protocol: the first time they enter the room when the light is off, they turn it on; on all other occasions, they do nothing. The counter follows a different protocol. The first $n - 2$ times that the light is on when he enters the interrogation room, he turns it off. Then the next time he enters the room when the light is on, he (truthfully) announces that everybody has been interrogated.[5]

In order to model this as a dynamic multi-agent system, we need to provide an initial epistemic model and the updates that are possible in that model. The n prisoners are named $0, \dots, n-1$. Prisoner 0 is the counter. In the protocol above we are only interested in the information the counter has. It is therefore sufficient only to model his information, and we can model this as a single-agent system. Atomic proposition p stands for for 'the light is on', and atomic propositions q_i, for $1 < i \leq n-1$, for 'prisoner i has turned on the light'. Formula $\bigwedge_{1<i\leq n-1} q_i$ is true when all prisoners except the counter have been interrogated.

We first define the update model that describes an event that a prisoner is interrogated. The ordinary prisoners execute the protocol that they turn on the light, if it is off and if the prisoner has not turned it on before. This event e_i can be seen as the simultaneous assignment.

$$p := q_i \rightarrow p \text{ and } \ q_i := p \rightarrow q_i$$

The counter executes a different protocol. He never turns on the light. He turns it off if it is on, otherwise he leaves it alone. This can be modelled as event e_0^p defined as

$$\text{if } \ p \ \text{ then } \ p := \bot$$

references before 2001. William Wu [29] mentions hearing about the riddle in 2001 and cites an IBM Research site `http://domino.watson.ibm.com/Comm/wwwr_ponder.nsf/challenges/July2002.html` where it is mentioned (in a 23 prisoner version) "This puzzle has been making the rounds of Hungarian mathematicians' parties." We did not find formal references. We thank Moshe Vardi for his advice.

[5]The riddle can also be solved when it is not known if the light is initially on or off. In that case the solution is for the non-counting prisoners to turn light on *twice* only if it is off, and the counter announces that everybody has been interrogated after he has turned off the light $2n - 2$ times.

and event $\mathsf{e}_0^{\neg p}$ consisting of precondition $\neg p$ only. In the single agent setting it is meaningless to model the announcement of the counter, and it is sufficient to show that eventually the precondition of that announcement is satisfied, namely that he eventually knows that all the other prisoners have turned on the light once. Finally, event $\mathsf{e}_\emptyset$ models that nothing happens when alternatively a prisoner could have been interrogated; it is therefore a 'skip' action. An overview of the different events is

$\mathsf{e}_\emptyset$	skip	'skip' means 'if $\top$ then $\emptyset$'
e_i	$p := q_i \rightarrow p$ and $q_i := p \rightarrow q_i$	for each $i > 0$
e_0^p	if p then $p := \bot$	
$\mathsf{e}_0^{\neg p}$	$\neg p$	

Agent 0 cannot distinguish between events e_i and $\mathsf{e}_\emptyset$, but e_0^p and $\mathsf{e}_0^{\neg p}$ that involve himself are of course distinguishable from those and from one another. The update model for the case of three prisoners is given is Figure 3.

We proceed with describing the epistemic model for the problem. The states in the model for the problem are characterized by the facts p and q_i true there. Although we defined the execution of an event as a transition from one epistemic state to another epistemic state, a different perspective is to see this as a 'shift' between points in a single (larger) epistemic model, i.e., more as a 'run' through a system. This allows for a simpler visualization. For the case of three prisoners, see Figure 4. Let us take a close look at this picture.

Initially all prisoners are in the dining area so there is still common knowledge that nobody has been interrogated: a singleton model with state $(\neg p, \neg q_1, \neg q_2)$ models that. This is the top state in the picture. The event that the prisoners leave the dining area is modelled as the non-deterministic execution of four possible events, namely all events with precondition $\neg p$: e_1, e_2, $\mathsf{e}_0^{\neg p}$, and $\mathsf{e}_\emptyset$. The last represents that even when noone has been interrogated yet, the counter considers it possible that someone has been interrogated. That this action is necessary becomes clear in case, later, the counter is interrogated himself and finds the light still off. He will then learn that noone was interrogated yet: a transition back to the top state of the model. Without the $\mathsf{e}_\emptyset$-event we could not have modelled this. When one of the other prisoners is interrogated, this leads to the light being on. No information change occurs as long as the counter is not interrogated. When this happens, he turns the light off and learns that at least one prisoner has been interrogated. The counter considers it possible that prisoner 1 has turned on the light and that prisoner 2 has turned on the light. Upon leaving the interrogation room, again the counter immediately considers it possible that other prisoners may have been interrogated, modelled by another informative $\mathsf{e}_\emptyset$ transition. This possibility can again be excluded by observing that the light is still off during yet another interrogation. Eventually the other prisoner is also interrogated, which leads to the light being on again and by observing this the counter learns that both prisoners have turned on the light once. This is modelled by another e_0^p-transition, now to the bottom state of the picture. There, the counter indeed knows that everyone has been interrogated (and in principle he can make an announcement to that effect).

It would be nice to also investigate what happens if we do take the information of the non-counting prisoners into account. Suppose for example that one such agent is always interrogated before the counter. That would mean the prisoner gets the same information as the counter, just a bit quicker. Indeed, this prisoner could announce that everyone has been interrogated before the counter. We leave this matter for the future.

6 Further research

Currently we consider several directions for further research.

An unresolved issue is whether updates can be described as compositions of 'purely epistemic' events (preconditions only) and 'purely ontic' events (postconditions only). In [11] it is shown for public events, for a different logical (more PDL-like) setting. Such a result would be in the line of our other 'normalization' results, and facilitate comparison to related approaches.

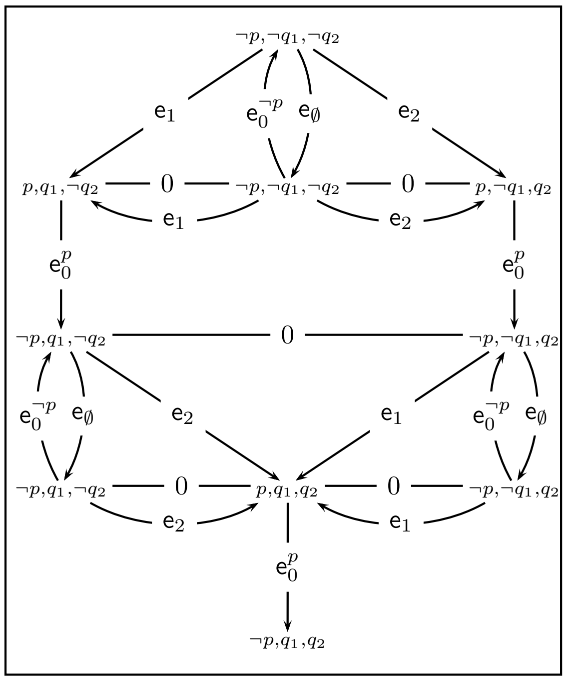


Figure 4: The initial epistemic state and the results of executing the update model for all possible events are pictured for the case of *three* prisoners trying to escape. The states are indicated by an atomic description. We assume reflexivity and transitivity of access for agent 0. We also did not draw the reflexive arrows for events that can be executed but do not yield information change. For example, in the top-left state events e_1, e_2, and $\mathsf{e}_\emptyset$ can also be executed but have no effect. In the bottom state agent 0 knows that all other prisoners have turned on the light at least once.

As said, the logic can be 'applied' to describe cryptographic bit exchange protocols, including protocols where keys change hands or are being sent between agents. The logic is very suitable for the description of protocols for computationally unlimited agents, such as described in the cited [9, 18]. Using dynamic logics may be an advantage given the availability of model checking tools for such logics, as e.g. the very versatile epistemic model checker DEMO [28]—van Eijck considers an extension of that model checker with assignments (postconditions): if that were the case, the match with the logic in this contribution would be perfect.

We intend to model the 'prisoners' example for the case where the state of the light is unknown at the beginning, and in a proper multi-agent setting that allows for more efficient protocols benefiting from non-counting prisoners' knowledge in exceptional situations.

The results for 'arbitrary belief change' suggest yet another possibly promising direction. Under certain conditions arbitrary formulas are realizable. What formulas are still realizable if one restricts the events to those considered suitable for specific problem areas, such as forms of multi-agent planning? And given a desirable formula (a 'postcondition' in another sense of the word), what are the initial conditions such that a sequence of events realizes it? This is the relation to AI problems concerning regression as pointed out in the introductory section, on which we think fast progress can now be made.

References

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] A. Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54(1):1–45, 2002.

[3] A. Baltag and L.S. Moss. Logics for epistemic programs. *Synthese*, 139:165–224, 2004. Knowledge, Rationality & Action 1–60.

[4] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. Technical report, Centrum voor Wiskunde en Informatica, Amsterdam, 1999. CWI Report SEN-R9922.

[5] A. Baltag, H.P. van Ditmarsch, and L.S. Moss. Epistemic logic and information update. In J.F.A.K. van Benthem and P. Adriaans, editors, *Handbook on the Philosophy of Information*, Amsterdam, 2006. Elsevier. In progress.

[6] J. Barwise and L.S. Moss. *Vicious Circles*. CSLI Publications, Stanford, 1996.

[7] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001. Cambridge Tracts in Theoretical Computer Science 53.

[8] S. Druiven. Knowledge development in games of imperfect information. Master's thesis, University of Groningen, Groningen, the Netherlands, 2002.

[9] M.J. Fischer and R.N. Wright. Bounds on secret key exchange using a random deal of cards. *Journal of Cryptology*, 9(2):71–99, 1996.

[10] A. Herzig, J. Lang, and P. Marquis. Action progression and revision in multiagent belief structures. In *Sixth Workshop on Nonmonotonic Reasoning, Action, and Change (NRAC 2005)*, 2005. `http://www.cse.unsw.edu.au/~nrac05/`.

[11] A. Herzig and T. De Lima. Epistemic actions and ontic actions: A unified logical framework. In J.S. Sichman et al., editors, *IBERAMIA-SBIA 2006, LNAI 4140*, pages 409–418. Springer, 2006.

[12] H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 387–394, 1991.

[13] B.P. Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 2006. To appear.

[14] N. Laverny. *Révision, mises à jour et planification en logique doxastique graduelle*. PhD thesis, Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France, 2006.

[15] Pit game rules. See `www.hasbro.com/common/instruct/pit.pdf`.

[16] J.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.

[17] G.R. Renardel de Lavalette. Changing modalities. *Journal of Logic and Computation*, 14(2):253–278, 2004.

[18] A. Stiglic. Computations with a deck of cards. *Theoretical Computer Science*, 259(1–2):671–678, 2001.

[19] J.F.A.K. van Benthem. Semantic parallels in natural language and computation. In *Logic Colloquium '87*, Amsterdam, 1989. North-Holland.

[20] J.F.A.K. van Benthem. Dynamic odds and ends. Technical report, University of Amsterdam, 1998. ILLC Research Report ML-1998-08.

[21] J.F.A.K. van Benthem, J. van Eijck, and B.P. Kooi. Logics of communication and change, 2006. Accepted for *Information and Computation*.

[22] H.P. van Ditmarsch. Descriptions of game actions. *Journal of Logic, Language and Information*, 11:349–365, 2002.

[23] H.P. van Ditmarsch. Some game theory of pit. In C. Zhang, H.W. Guesgen, and W.K. Yeap, editors, *Proceedings of PRICAI 2004 (Eighth Pacific Rim International Conference on Artificial Intelligence)*, pages 946–947. Springer, 2004. LNAI 3157.

[24] H.P. van Ditmarsch. Belief change and dynamic logic. In J. Delgrande, J. Lang, H. Rott, and J.-M. Tallon, editors, *Belief Change in Rational Agents: Perspectives from Artificial Intelligence, Philosophy, and Economics*, number 05321 in Dagstuhl Seminar Proceedings, Dagstuhl, 2005. Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl. `http://drops.dagstuhl.de/opus/volltexte/2005/337`.

[25] H.P. van Ditmarsch. The logic of pit. *Knowledge, Rationality & Action (Synthese)*, 149(2):343–375, 2006.

[26] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Dynamic epistemic logic with assignment. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 05)*, volume 1, pages 141–148, New York, 2005. ACM Inc.

[27] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Dynamic epistemic logic. Manuscript, 2006.

[28] J. van Eijck. Dynamic epistemic modelling. Technical report, Centrum voor Wiskunde en Informatica, Amsterdam, 2004. CWI Report SEN-E0424.

[29] W. Wu. 100 prisoners and a lightbulb. `http://www.ocf.berkeley.edu/~wwu/papers/100prisonersLightBulb.pdf`, 2001.

Appendix: Completeness

As usual, the completeness proof consists of building a model for an given formula, thus showing that all consistent formulas are satisfiable, and by contraposition completeness of the logic. We first define a canonical model for a given formula (Definition 22, below), then prove a number of relevant properties of the model (Lemma 21, on page 22), from which the for completeness required 'Truth Lemma' then follows without great difficulty (Lemma 28, on page 28).

Let $Sub(\varphi)$ be the set of subformulas of φ.

Definition 19 (closure) Let $cl : \mathscr{L} \to \wp(\mathscr{L})$, such that for every $\varphi \in \mathscr{L}$, $cl(\varphi)$ is the smallest set such that:

1. $\varphi \in cl(\varphi)$,
2. if $\psi \in cl(\varphi)$, then $Sub(\psi) \subseteq cl(\varphi)$
3. if $\psi \in cl(\varphi)$ and ψ is not a negation, then $\neg\psi \in cl(\varphi)$
4. if $[B^+]\psi \in cl(\varphi)$, then $\{[a][B^+]\psi \mid a \in B\} \subseteq cl(\varphi)$
5. if $[\mathsf{U}, \mathsf{e}]p \in cl(\varphi)$, then $(\mathsf{pre}(\mathsf{e}) \to \mathsf{post}_{\mathsf{e}}(p)) \in cl(\varphi)$
6. if $[\mathsf{U}, \mathsf{e}]\neg\psi \in cl(\varphi)$, then $(\mathsf{pre}(\mathsf{e}) \to \neg[\mathsf{U}, \mathsf{e}]\psi) \in cl(\varphi)$
7. if $[\mathsf{U}, \mathsf{e}](\psi \wedge \chi) \in cl(\varphi)$, then $([\mathsf{U}, \mathsf{e}]\psi \wedge [\mathsf{U}, \mathsf{e}]\chi) \in cl(\varphi)$
8. if $[\mathsf{U}, \mathsf{e}][a]\psi \in cl(\varphi)$ and $(\mathsf{e}, \mathsf{f}) \in \mathsf{R}(a)$, then $(\mathsf{pre}(\mathsf{e}) \to [a][\mathsf{U}, \mathsf{f}]\psi) \in cl(\varphi)$
9. if $[\mathsf{U}, \mathsf{e}][B^+]\psi \in cl(\varphi)$, then $\{[\mathsf{U}, \mathsf{f}]\psi \mid (\mathsf{e}, \mathsf{f}) \in \mathsf{R}(B)^+\} \subseteq cl(\varphi)$ and $\{[a][\mathsf{U}, \mathsf{f}][B^+]\psi \mid a \in B \text{ and } (\mathsf{e}, \mathsf{f}) \in \mathsf{R}(B)^+\} \subseteq cl(\varphi)$
10. if $[\mathsf{U}, \mathsf{e}][\mathsf{U}', \mathsf{e}']\psi \in cl(\varphi)$, then $[(\mathsf{U}, \mathsf{e}); (\mathsf{U}', \mathsf{e}')]\psi \in cl(\varphi)$ ⊣

Lemma 20 $cl(\varphi)$ is finite for all formulas $\varphi \in \mathscr{L}$. ⊣

Proof Straightforward by induction on φ.

Definition 21 (maximal consistent in Φ) Let $\Phi \subseteq \mathscr{L}$ be the closure of some formula. Γ is maximal consistent in Φ iff

1. $\Gamma \subseteq \Phi$
2. Γ is consistent: $\Gamma \nvdash \bot$
3. Γ is maximal in Φ: there is no $\Gamma' \subseteq \Phi$ such that $\Gamma \subset \Gamma'$ and $\Gamma' \nvdash \bot$.

Let $\underline{\Gamma} = \bigwedge\{\varphi \mid \varphi \in \Gamma\}$. ⊣

Definition 22 (canonical model for Φ) Let Φ be the closure of some formula. The canonical model $M^c = (S^c, R^c, V^c)$ for Φ is defined as follows

- $S^c = \{\Gamma \mid \Gamma \text{ is maximal consistent in } \Phi\}$
- $(\Gamma, \Delta) \in R^c(a)$ iff for all φ if $[a]\varphi \in \Gamma$, then $\varphi \in \Delta\}$
- $V^c(p) = \{\Gamma \in S^c \mid p \in \Gamma\}$ ⊣

Lemma 23 (Lindenbaum) Let Φ be the closure of some formula. Every consistent subset of Φ is a subset of a maximal consistent set in Φ. ⊣

Proof The proof is trivial since Φ is finite by Lemma 20.

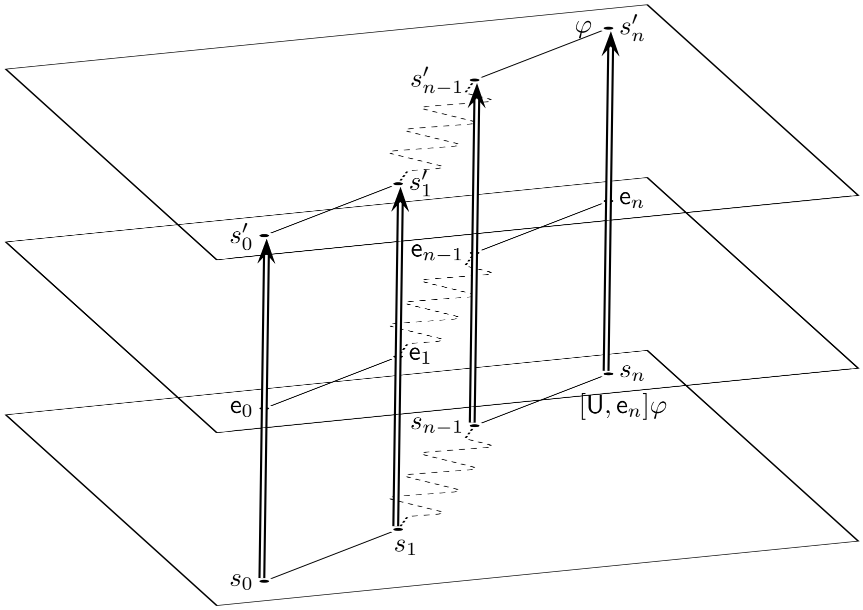


Figure 5: This Figure helps to make the semantic intuition for formulas of the form $[\mathsf{U}, \mathsf{e}][B^+]\varphi$ clearer. The model below is the original model. The model in the middle is the update model. The top model is the resulting model.

Semantically it is quite clear what it means for a formula of the form $[\mathsf{U}, \mathsf{e}][B^+]\varphi$ to be true in a state s. It means that every B-path from (s, e) in the resulting model ends in a φ-state. But that means that in the original model, a B-path for which there is a matching B-path $\mathsf{e}_0, \ldots \mathsf{e}_n$ in the update model ends in a $[\mathsf{U}, \mathsf{e}_n]\varphi$-state. This is shown in Figure 5. In view of this, in the proofs we will use the notion of a $B\mathsf{Uef}$-path.

Definition 24 (paths) A B-path from Γ is a sequence $\Gamma_0, \ldots, \Gamma_n$ of maximal consistent sets in Φ such that for all k (where $0 \leq k < n$) there is an agent $a \in B$ such that $(\Gamma_k, \Gamma_{k+1}) \in R^c(a)$ and $\Gamma_0 = \Gamma$.

A $B\mathsf{Uef}$-path from Γ is a B-path $\Gamma_0, \ldots \Gamma_n$ from Γ such that there is a B-path $\mathsf{e}_0, \ldots \mathsf{e}_n$ from e to f in U and for all $k < n$ there is an agent a such that $(\Gamma_k, \Gamma_{k+1}) \in R^c(a)$ and $(\mathsf{e}_k, \mathsf{e}_{k+1}) \in \mathsf{R}(a)$ and for all $k \leq n$ it is the case that $\mathsf{pre}(\mathsf{e}_k) \in \Gamma_k$. ⊣

Lemma 25 Let Φ be the closure of some formula. Let $M^c = (S^c, R^c, V^c)$ be the canonical model for Φ. If Γ and Δ are maximal consistent sets in Φ, then

1. Γ is deductively closed in Φ;
2. $\vdash \bigvee\{\underline{\Gamma} \mid \Gamma \in S^c\}$;
3. if $\neg\varphi \in \Phi$, then $\varphi \in \Gamma$ iff $\neg\varphi \notin \Gamma$;
4. if $(\varphi \wedge \psi) \in \Phi$, then $(\varphi \wedge \psi) \in \Gamma$ iff $\varphi \in \Gamma$ and $\psi \in \Gamma$;
5. if $\underline{\Gamma} \wedge \langle a \rangle \underline{\Delta}$ is consistent, then $(\Gamma, \Delta) \in R^c(a)$;
6. if $[B^+]\varphi \in \Phi$, then $[B^+]\varphi \in \Gamma$ iff every B-path from Γ is a φ-path;
7. if $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Phi$, then $[\mathsf{U}, \mathsf{e}][B^+]\psi \in \Gamma$ iff for all $\mathsf{f} \in \mathsf{E}$ every $B\mathsf{Uef}$-path from Γ ends in an $[\mathsf{U}, \mathsf{f}]\psi$-state.

Proof

1. Suppose $\varphi \in \Phi$ and $\vdash \underline{\Gamma} \to \varphi$. Then $\Gamma \cup \{\varphi\}$ is also consistent in Φ. Therefore, by maximality, $\varphi \in \Gamma$.

2. Suppose that $\bigvee\{\underline{\Gamma} \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$ is not a tautology, i.e. that $\neg \bigvee\{\underline{\Gamma} \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$ is consistent. Therefore $\bigwedge\{\neg\underline{\Gamma} \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$ is consistent. Therefore $\bigwedge\{\bigvee\{\neg\varphi \mid \varphi \in \Gamma\} \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$ is consistent. Therefore for every maximal consistent Γ in $cl(\varphi)$ there is a formula $\varphi_\Gamma \in \Gamma$ such that $\{\neg\varphi_\Gamma \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$ is consistent. Therefore by the Lindenbaum Lemma (and the law of double negation) there is a maximal consistent set in $cl(\varphi)$ that is inconsistent with every maximal consistent set in $cl(\varphi)$. That is a contradiction. Therefore $\vdash \bigvee\{\underline{\Gamma} \mid \Gamma$ is maximal consistent in $cl(\varphi)\}$.

3. Suppose $\neg\varphi \in \Phi$.

   From left to right. Suppose $\varphi \in \Gamma$. By consistency $\neg\varphi \notin \Gamma$.

   From right to left. Suppose $\neg\varphi \notin \Gamma$. By maximality $\vdash (\underline{\Gamma} \wedge \neg\varphi) \to \bot$. But then $\vdash \underline{\Gamma} \to \varphi$. Therefore $\varphi \in \Gamma$, because Γ is deductively closed in Φ.

4. Suppose $(\varphi \wedge \psi) \in \Gamma$. Then $(\varphi \wedge \psi) \in \Gamma$ is equivalent to $\varphi \in \Gamma$ and $\psi \in \Gamma$ because Γ is deductively closed in Φ.

5. Suppose that $\underline{\Gamma} \wedge \langle a \rangle \underline{\Delta}$ is consistent. Suppose that it is not the case that $(\Gamma, \Delta) \in R^c(a)$. Therefore there is a formula φ such that

   $$[a]\varphi \in \Gamma \text{ but } \varphi \notin \Delta$$

   By 2 of this lemma $\neg\varphi \in \Delta$ (or if φ is a negation $\neg\chi$ then $\chi \in \Delta$). But then, $\vdash \langle a \rangle \underline{\Delta} \to \langle a \rangle \neg\varphi$. Since $[a]\varphi \in \Gamma$, it must be the case that $\underline{\Gamma} \wedge \langle a \rangle \underline{\Delta}$ is inconsistent, contradicting our initial assumption. Therefore $(\Gamma, \Delta) \in R^c(a)$.

6. From left to right. We proceed by induction on the length of the B-path, but we load the induction (we prove something stronger). We show that if $[B^+]\varphi \in \Gamma$, then every B-path is a φ-path *and* a $[B^+]\varphi$-path. Suppose that $[B^+]\varphi \in \Gamma$.

   Base case Suppose the length of the path is 1, i.e. the path consists of two sets Γ_0 and Γ_1. Therefore there is some agent in $a \in B$ such that $(\Gamma_0, \Gamma_1) \in R^c(a)$. Since $\vdash [B^+]\varphi \to [a](\varphi \wedge [B^+]\varphi)$ and Γ_1 is deductively closed in Φ, it must be the case that both $\varphi \in \Gamma_1$ and $[B^+]\varphi \in \Gamma_1$.

   Induction hypothesis Suppose that if $[B^+]\varphi \in \Gamma$, then every B-path of length n is a φ-path and a $[B^+]\varphi$-path.

   Induction step Take a B-path of length $n+1$ from Γ. By the induction hypothesis $[B^+]\varphi \in \Gamma_n$. Now we can proceed in the same way that we did in the base case and conclude that $\varphi \in \Gamma_{n+1}$ and $[B^+]\varphi \in \Gamma_{n+1}$.

   From right to left. Suppose that every B-path from Γ is a φ-path. Let $S_{B,\varphi}$ be the set of all maximal consistent sets Δ in Φ such that every B-path from Δ is a φ-path. Now consider the formula

   $$\chi = \bigvee_{\Delta \in S_{B,\varphi}} \underline{\Delta}$$

   We prove the following

   $$\Gamma \vdash \chi \tag{a}$$
   $$\vdash \chi \to \varphi \tag{b}$$
   $$\vdash \chi \to [B]\chi \tag{c}$$

From these it follows that $[B^+]\varphi \in \Gamma$. Because from (c) it follows by necessitation that $\vdash [B^+](\chi \to [B]\chi)$. By the induction axiom this implies that $\vdash \chi \to [B^+]\chi$. By (a) this implies that $\Gamma \vdash [B^+]\chi$. By (b) and necessitation and distribution for $[B^+]$ this implies that $\Gamma \vdash [B^+]\varphi$. Therefore $[B^+]\varphi \in \Gamma$.

(a) $\Gamma \vdash \chi$, because $\underline{\Gamma}$ is one of the disjuncts of χ.

(b) Note that $\varphi \in \Delta$ if Δ is in $S_{B,\varphi}$. Therefore φ is a conjunct of every disjunct of χ. Therefore $\vdash \chi \to \varphi$.

(c) Suppose toward a contradiction that $\chi \wedge \neg[B]\chi$ is consistent. Since χ is a disjunction there must be a disjunct $\underline{\Delta}$ such that $\underline{\Delta} \wedge \neg[B]\chi$ is consistent. By similar reasoning there must be an $a \in B$ such that $\underline{\Delta} \wedge \langle a \rangle \neg\chi$ is consistent. Since $\vdash \bigvee\{\underline{\Delta} \mid \Delta \in S^c\}$ (by item 2 of this lemma), it is the case that $\underline{\Delta} \wedge \langle a \rangle \bigvee_{\Theta \notin S_{B,\varphi}} \underline{\Theta}$ is consistent. Then, by modal reasoning, $\underline{\Delta} \wedge \bigvee_{\Theta \notin S_{B,\varphi}} \langle a \rangle \underline{\Theta}$ is consistent. Therefore there must be a $\Theta \notin S_{B,\varphi}$ which is maximal consistent in Φ, such that $\underline{\Delta} \wedge \langle a \rangle \underline{\Theta}$ is consistent. But then by (4) of this Lemma, $(\Delta, \Theta) \in R^c(a)$. Since Θ is not in $S_{B,\varphi}$, there must be a B-path from Θ which is not a φ-path. But then there is a B-path from Δ which is not a φ-path. This contradicts that $\Delta \in S_{B,\varphi}$, which it is because $\underline{\Delta}$ is one of χ's disjuncts. Therefore $\vdash \chi \to [B]\chi$.

7. From left to right. Suppose $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Gamma$. We now continue by induction on the length of the path, but we load the induction. We show that if $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Gamma$, then every $B\mathsf{Uef}$-path ends in an $[\mathsf{U}, \mathsf{f}]\varphi$-state which is also an $[\mathsf{U}, \mathsf{f}][B^+]\varphi$-state.

Base case Suppose the length of the path is 0. Therefore we merely need to show that if $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Gamma_0$, then $[\mathsf{U}, \mathsf{e}]\varphi \in \Gamma_0$. This can be shown as follows. Observe that $[B^+]\varphi \to \varphi$. By necessitation and distribution for $[\mathsf{U}, \mathsf{e}]$ we have that $[\mathsf{U}, \mathsf{e}][B^+]\varphi \to [\mathsf{U}, \mathsf{e}]\varphi$. Therefore $[\mathsf{U}, \mathsf{e}]\varphi \in \Gamma_0$.

Induction hypothesis Suppose that every $B\mathsf{Uef}$-path of length n ends in an $[\mathsf{U}, \mathsf{f}]\psi$-state which is also an $[\mathsf{U}, \mathsf{f}][B^+]$-state.

Induction step Suppose now that the path is of length $n+1$, i.e. there is a $B\mathsf{Uef}$-path $\Gamma_0, \ldots \Gamma_n, \Gamma_{n+1}$. Therefore there is a $B\mathsf{Ueg}$-path for some g from Γ_0 to Γ_n. By the induction hypothesis we may assume that $[\mathsf{U}, \mathsf{g}][B^+]\varphi \in \Gamma_n$. Since there is a $B\mathsf{Uef}$-path, there is an agent a such that $(\Gamma_n, \Gamma_{n+1}) \in R^c(a)$ and $(\mathsf{g}, \mathsf{f}) \in \mathsf{R}(a)$. By the mix axiom we have that $\vdash [B^+]\psi \to [a][B^+]\psi$. By applying necessitation and distribution for $[\mathsf{U}, \mathsf{g}]$ we get $\vdash [\mathsf{U}, \mathsf{g}][B^+]\varphi \to [\mathsf{U}, \mathsf{g}][a][B^+]\varphi$. By the axiom for updates and knowledge $\vdash [\mathsf{U}, \mathsf{g}][a][B^+]\varphi \to (\mathsf{pre}(\mathsf{g}) \to [a][\mathsf{U}, \mathsf{f}][B^+]\varphi)$. Since $\mathsf{pre}(\mathsf{g}) \in \Gamma_n$, it must be the case that $\Gamma_n \vdash [a][\mathsf{U}, \mathsf{f}][B^+]\varphi$. Therefore by the definition of $R^c(a)$, it is the case that $[a][\mathsf{U}, \mathsf{f}][B^+]\varphi \in \Gamma_{n+1}$, and so $[\mathsf{U}, \mathsf{f}][B^+]\psi \in \Gamma_{n+1}$. And, by similar reasoning as in the base case, $[\mathsf{U}, \mathsf{f}]\psi \in \Gamma_{n+1}$.

From right to left. Suppose that every $B\mathsf{Uef}$-path from Γ ends in an $[\mathsf{U}, \mathsf{f}]\varphi$-state. Let $S_{B,\mathsf{U},\mathsf{f},\varphi}$ be the set of maximal consistent sets Δ such that every $B\mathsf{Ufg}$-path from Δ ends in an $[\mathsf{U}, \mathsf{g}]\varphi$-state for all g such that $(\mathsf{f}, \mathsf{g}) \in \mathsf{R}(B)^+$. Now consider the formulas

$$\chi_{\mathsf{f}} = \bigvee_{\Delta \in S_{B,\mathsf{U},\mathsf{f},\varphi}} \underline{\Delta}$$

We will show the following

$$\Gamma \vdash \chi_{\mathsf{e}} \quad \text{(a)}$$

$$\vdash \chi_{\mathsf{f}} \to [\mathsf{U}, \mathsf{f}]\varphi \quad \text{(b)}$$

$$\vdash (\chi_{\mathsf{f}} \wedge \mathsf{pre}(\mathsf{f})) \to [a]\chi_{\mathsf{g}} \text{ if } (\mathsf{f}, \mathsf{g}) \in \mathsf{R}(a) \quad \text{(c)}$$

From these it follows that $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Gamma$. By applying the updates and common knowledge rule to (b) and (c) it follows that $\vdash \chi_{\mathsf{e}} \to [\mathsf{U}, \mathsf{e}][B^+]\psi$. By (a) it follows that $\Gamma \vdash [\mathsf{U}, \mathsf{e}][B^+]\varphi$. Therefore $[\mathsf{U}, \mathsf{e}][B^+]\varphi \in \Gamma$.

(a) This follows immediately, because $\Gamma \in S_{B,\mathsf{U},\mathsf{e},\varphi}$, and $\underline{\Gamma}$ is one of χ_{e}'s disjuncts.

(b) Since every $B\mathsf{Ufg}\varphi$-path ends in an $[\mathsf{U},\mathsf{g}]\varphi$-state for every $\Delta \in S_{B,\mathsf{U},\mathsf{f},\varphi}$ it must be the case that the length one $B\mathsf{Uff}\varphi$-path ends in an $[\mathsf{U},\mathsf{f}]\varphi$-state. Therefore $[\mathsf{U},\mathsf{f}]\varphi \in \Delta$. Therefore, $[\mathsf{U},\mathsf{f}]\varphi$ is one of the conjuncts in every disjunct of χ_{f}. Therefore by propositional reasoning $\chi \rightarrow [\mathsf{U},\mathsf{f}]\varphi$.

(c) Suppose toward a contradiction that $\chi_{\mathsf{f}} \wedge \mathsf{pre}(\mathsf{f}) \wedge \neg[a]\chi_{\mathsf{g}}$ is consistent. Because χ_{f} is a disjunction there must be a disjunct $\underline{\Theta}$ such that $\underline{\Theta} \wedge \mathsf{pre}(\mathsf{f}) \wedge \neg[a]\chi_{\mathsf{g}}$ is consistent. Since $\underline{\Theta} \wedge \mathsf{pre}(\mathsf{f})$ is consistent in Φ and $\mathsf{pre}(\mathsf{f}) \in \Phi$ it must be the case that $\mathsf{pre}(\mathsf{f}) \in \Theta$. So $\underline{\Theta} \wedge \langle a\rangle\neg\chi_{\mathsf{g}}$ is consistent. Since $\vdash \bigvee\{\underline{\Gamma} \mid \Gamma \in S^c\}$, there must be a Ξ in the complement of $S_{B,\mathsf{U},\mathsf{g},\varphi}$ such that $\underline{\Theta} \wedge \langle a\rangle\underline{\Xi}$ is consistent. By item 5 of Lemma 25 this implies that $(\Theta, \Xi) \in R^c(a)$. But since Ξ is not in $S_{B,\mathsf{U},\mathsf{g},\varphi}$ there must be an h such that there is a $B\mathsf{Ugh}$-path not ending in an $[\mathsf{U},\mathsf{h}]\varphi$-state from Ξ. But then there is also a $B\mathsf{Ufh}$-path from Θ not ending in an $[\mathsf{U},\mathsf{h}]\varphi$-state. This contradicts that $\Theta \in S_{B,\mathsf{U},\mathsf{f},\varphi}$, but it must be, because $\underline{\Theta}$ is one of χ_{f}'s disjuncts. Therefore $\vdash (\chi_{\mathsf{f}} \wedge \mathsf{pre}(\mathsf{f})) \rightarrow [a]\chi_{\mathsf{g}}$.

Proofs by induction on formulas use the inductive definition of the logical language. In the inductive step one uses the induction hypothesis for subformulas of that formula. It seems however that this use of the induction hypothesis is not enough for the case at hand. For example $\neg[\mathsf{U},\mathsf{e}]\varphi$ is not a subformula of $[\mathsf{U},\mathsf{e}]\neg\varphi$, but we would like to apply the induction hypothesis to the former when we are proving the inductive step for the latter. Since the inductive definition of the logical language does not say anything about the order of these formulas we can impose an order on them that suits our purposes best. For this we define the following *complexity measure* on the language of *UM*.

Definition 26 (complexity) The complexity $c : \mathscr{L} \rightarrow \mathbb{N}$ is defined as follows:

$$\begin{array}{lcl} c(p) & = & 1 \\ c(\neg\varphi) & = & 1 + c(\varphi) \\ c(\varphi \wedge \psi) & = & 1 + \max(c(\varphi), c(\psi)) \\ c([a]\varphi) & = & 1 + c(\varphi) \\ c([B^+]\varphi) & = & 1 + c(\varphi) \\ c([\mathsf{U},\mathsf{e}]\varphi) & = & (4 + c(\mathsf{U})) \cdot c(\varphi) \end{array}$$

where $c(\mathsf{U}) = \max(\{c(\mathsf{pre}(\mathsf{e})) \mid \mathsf{e} \in \mathsf{U}\} \cup \{c(\varphi) \mid \varphi \in \mathrm{ran}(\mathsf{post}(\mathsf{e}))\})$. ⊣

We can safely take the maximum of the complexity of the preconditions and formulas in the range of substitutions in the action model, since action models are *finite*. Therefore the complexity of a formula or an action model will always be a natural number. In the definition the number 4 appears in the clause for update models. It seems arbitrary, but it is in fact the least natural number that gives us the following properties.

Lemma 27 For all φ, ψ, and χ:

1. $c(\psi) \geq c(\varphi)$ for all $\varphi \in Sub(\psi)$.
2. $c([\mathsf{U},\mathsf{e}]p) > c(\mathsf{pre}(\mathsf{e}) \rightarrow \mathsf{post}_{\mathsf{e}}(p))$
3. $c([\mathsf{U},\mathsf{e}]\neg\varphi) > c(\mathsf{pre}(\mathsf{e}) \rightarrow \neg[\mathsf{U},\mathsf{e}]\varphi)$
4. $c([\mathsf{U},\mathsf{e}](\varphi \wedge \psi) > c([\mathsf{U},\mathsf{e}]\varphi \wedge [\mathsf{U},\mathsf{e}]\varphi)$
5. $c([\mathsf{U},\mathsf{e}][a]\varphi) > c(\mathsf{pre}(\mathsf{e}) \rightarrow [a][\mathsf{U},\mathsf{f}]\varphi)$ for all $\mathsf{f} \in \mathsf{U}$
6. $c([\mathsf{U},\mathsf{e}][B^+]\varphi) > c([\mathsf{U},\mathsf{f}]\varphi)$ for all $\mathsf{f} \in \mathsf{U}$
7. $c([\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\varphi) > c([(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')]\varphi)$

Proof 1. By induction on ψ.

Base case If ψ is a propositional variable, its complexity is 1 and it is its only subformula.

Induction hypothesis $c(\psi) \geq c(\varphi)$ if $\varphi \in Sub(\psi)$ and $c(\chi) \geq c(\varphi)$ if $\varphi \in Sub(\xi)$.

Induction step We proceed by cases

negation Suppose that φ is a subformula of $\neg\psi$. Then φ is either $\neg\psi$ or a subformula of ψ. In the former case, we simply observe that the complexity of every formula is greater than or equal to its own complexity. In the latter case, the complexity of $\neg\psi$ equals $1 + c(\psi)$. Therefore if φ is a subformula of ψ it follows immediately from the induction hypothesis that $c(\psi) \geq c(\varphi)$.

conjunction Suppose that φ is a subformula of $\psi \wedge \chi$. Then φ is either $\psi \wedge \chi$ or it is a subformula of ψ or χ. Again in the former case, the complexity of every formula is greater than or equal to its own complexity. In the latter case the complexity of $\psi \wedge \chi$ equals $1 + \max(c(\varphi), c(\psi))$. Simple arithmetic and the induction hypothesis gives us that $c(\psi \wedge \chi) \geq c(\varphi)$.

individual epistemic operator This is completely analogous to the case for negation.

common knowledge This is also completely analogous to the case for negation.

update models Suppose that φ is a subformula of $[\mathsf{U}, \mathsf{e}]\psi$. Then φ is either $[\mathsf{U}, \mathsf{e}]\psi$ or it is a subformula of ψ or it is a precondition or postcondition in U. Again, in the former case, the complexity of every formula is greater than or equal to its own complexity. In the latter two cases, the complexity of $[\mathsf{U}, eone]\chi$ equals $(4 + c(\mathsf{U})) \cdot c(\chi)$. Simple arithmetic and the induction hypothesis gives us that $c([\mathsf{U}, \mathsf{e}]\chi) \geq c(\varphi)$.

2. $$\begin{aligned} c(\mathsf{pre}(\mathsf{e}) \rightarrow \mathsf{post}_\mathsf{e}(p)) &= c(\neg(\mathsf{pre}(\mathsf{e}) \wedge \neg\mathsf{post}_\mathsf{e}(p))) \\ &= 1 + c(\mathsf{pre}(\mathsf{e}) \wedge \neg\mathsf{post}_\mathsf{e}(p)) \\ &= 2 + \max(c(\mathsf{pre}(\mathsf{e})), 1 + c(\mathsf{post}_\mathsf{e}(p)) \end{aligned}$$

   and

   $$\begin{aligned} c([\mathsf{U}, \mathsf{e}]p) &= (4 + c(\mathsf{U})) \cdot 1 \\ &= 4 + c(\mathsf{U}) \end{aligned}$$

   and

   $c(\mathsf{U}) \geq \max(c(\mathsf{pre}(\mathsf{e})), c(\mathsf{post}_\mathsf{e}(p)))$

   From this it is clear that $c([\mathsf{U}, \mathsf{e}]p) > c(\mathsf{pre}(\mathsf{e}) \rightarrow \mathsf{post}_\mathsf{e}(p))$

3. $$\begin{aligned} c([\mathsf{U}, \mathsf{e}]\neg\varphi) &= (4 + c(\mathsf{U})) \cdot (1 + c(\varphi)) \\ &= 4 + c(\mathsf{U}) + 4 \cdot c(\varphi) + c(\mathsf{U}) \cdot c(\varphi) \end{aligned}$$

   and

   $$\begin{aligned} c(\mathsf{pre}(\mathsf{e}) \rightarrow \neg[\mathsf{U}]\varphi) &= c(\neg(\mathsf{pre}(\mathsf{e}) \wedge \neg\neg[\mathsf{U}]\varphi)) \\ &= 2 + \max(c(\mathsf{pre}(\mathsf{e}), 2 + ((4 + c(\mathsf{U})) \cdot c(\varphi))) \\ &= 2 + \max(c(\mathsf{pre}(\mathsf{e}), 2 + 4 \cdot c(\psi) + c(\varphi) \cdot c(\psi)) \end{aligned}$$

   The latter equals $2 + c(\varphi)$ or $4 + 4 \cdot c(\psi) + c(\varphi) \cdot c(\psi)$. Both are less than $4 + c(\varphi) + 4 \cdot c(\psi) + c(\varphi) \cdot c(\psi)$.

4. Assume, without loss of generality, that $c(\varphi) \geq c(\psi)$. Then

   $$\begin{aligned} c([\mathsf{U}, \mathsf{e}](\varphi \wedge \psi)) &= (4 + c(\mathsf{U})) \cdot (1 + \max(c(\varphi), c(\psi))) \\ &= 4 + c(\mathsf{U}) + 4 \cdot \max(c(\varphi), c(\psi)) + c(\mathsf{U}) \cdot \max(c(\varphi), c(\psi)) \\ &= 4 + c(\mathsf{U}) + 4 \cdot c(\varphi) + c(\mathsf{U}) \cdot c(\varphi) \end{aligned}$$

   and

   $$\begin{aligned} c([\mathsf{U}, \mathsf{e}]\varphi \wedge [\mathsf{U}, \mathsf{e}]\psi) &= 1 + \max((4 + c(\mathsf{U})) \cdot c(\varphi), (4 + c(\mathsf{U})) \cdot c(\psi)) \\ &= 1 + ((4 + c(\mathsf{U})) \cdot c(\varphi)) \\ &= 1 + 4 \cdot c(\varphi) + c(\mathsf{U}) \cdot c(\varphi) \end{aligned}$$

   The latter is less than the former

5. This case is completely analogous to the case for negation.

6. This case is straightforward.

7. $c([\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\varphi) = (4+c(\mathsf{U}))\cdot(4+c(\mathsf{U}'))\cdot c(\varphi)$
   $= (16+4\cdot c(\mathsf{U})+4\cdot c(\mathsf{U}')+c(\mathsf{U})\cdot c(\mathsf{U}'))\cdot c(\varphi)$

   and

   $c([(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')]\varphi = (4+c((\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')))\cdot c(\varphi)$

   So now the inequality holds iff

   $$(16+4\cdot c(\mathsf{U})+4\cdot c(\mathsf{U}')+c(\mathsf{U})\cdot c(\mathsf{U}'))\cdot c(\varphi)c(\mathsf{U})\cdot c(\mathsf{U}')) > (4+c((\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')))$$

   The complexity of $(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')$ is at most the following.

   $\max(\{c(\mathsf{pre}(\mathsf{f})\wedge[\mathsf{U},\mathsf{f}]\mathsf{pre}(\mathsf{f}')) \mid \mathsf{f}\in\mathsf{U} \text{ and } \mathsf{f}'\in\mathsf{U}'\}\cup$
   $\{c([\mathsf{U},\mathsf{f}]\mathsf{post}'_{\mathsf{f}'}(p)) \mid \mathsf{f}\in\mathsf{U},\mathsf{f}'\in\mathsf{U}' \text{ and } p\in dom(\mathsf{post}'(\mathsf{f}'))\})$

   Worst case this equals

   $1+\max(c(\mathsf{U}),(4+c(\mathsf{U}))\cdot c(\mathsf{U}')) = 1+((4+c(\mathsf{U}))\cdot c(\mathsf{U}'))$
   $= 5+4\cdot c(\mathsf{U}')+c(\mathsf{U})\cdot c(\mathsf{U}')$

   Therefore the inequality clearly holds.

Lemma 28 (truth) Let Φ be the closure for some formula. Let $M^c = (S^c, R^c, V^c)$ be the canonical model for Φ. For all $\Gamma\in S^c$, and all $\varphi\in\Phi$:

$$\varphi\in\Gamma \text{ iff } (M^c,\Gamma)\models\varphi$$

Proof Suppose $\varphi\in\Phi$. We now continue by induction on $c(\varphi)$.

Base case Suppose $c(\varphi)=1$. Therefore φ is a propositional variable p. This case now follows straightforwardly from the definition of the valuation in the canonical model.

Induction hypothesis For all φ such that $c(\varphi)\leq n$: $\varphi\in\Gamma$ iff $(M^c,\Gamma)\models\varphi$.

Induction step Suppose $c(\varphi)=n+1$. The cases for negation, conjunction, and individual epistemic operators follow straightforwardly from the induction hypothesis. We distinguish the following cases:

the case for $[B^+]\psi$: Suppose $[B^+]\psi\in\Gamma$. From Lemma 25, item 6 it follows that this is the case iff every B path from Γ is a ψ path. It follows from the induction hypothesis that this is equivalent to saying that ψ is true along these paths. Therefore this is equivalent to $(M^c,\Gamma)\models[B^+]\psi$

the case for $[\mathsf{U},\mathsf{e}]p$: Suppose $[\mathsf{U},\mathsf{e}]p\in\Gamma$. Given that $[\mathsf{U},\mathsf{e}]p\in\Phi$, $[\mathsf{U},\mathsf{e}]p\in\Gamma$ is equivalent to $\mathsf{pre}(\mathsf{e})\to\mathsf{post}_{\mathsf{e}}(p)\in\Gamma$ by the update and atoms axiom. By item 2 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent to $(M^c,\Gamma)\models\mathsf{pre}(\mathsf{e})\to\mathsf{post}_{\mathsf{e}}(p)$. By the semantics this is equivalent to $(M^c,\Gamma)\models[\mathsf{U},\mathsf{e}]p$.

the case for $[\mathsf{U},\mathsf{e}]\neg\psi$: Suppose $[\mathsf{U},\mathsf{e}]\neg\psi\in\Gamma$. Given that $[\mathsf{U},\mathsf{e}]\neg\psi\in\Phi$, $[\mathsf{U},\mathsf{e}]\neg\psi\in\Gamma$ is equivalent to $(\mathsf{pre}(\mathsf{e})\to\neg[\mathsf{U},\mathsf{e}]\psi)\in\Gamma$ by the update and negation axiom. By item 3 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent to $(M^c,\Gamma)\models\mathsf{pre}(\mathsf{e})\to\neg[\mathsf{U},\mathsf{e}]\psi$. By the semantics this is equivalent to $(M^c,\Gamma)\models[\mathsf{U},\mathsf{e}]\neg\psi$.

the case for $[\mathsf{U},\mathsf{e}](\psi\wedge\chi)$: Suppose $[\mathsf{U},\mathsf{e}](\psi\wedge\chi)\in\Gamma$. Given that $[\mathsf{U},\mathsf{e}](\psi\wedge\chi)\in\Phi$, $[\mathsf{U},\mathsf{e}](\psi\wedge\chi)\in\Gamma$ is equivalent to $([\mathsf{U},\mathsf{e}]\psi\wedge[\mathsf{U},\mathsf{e}]\chi)\in\Gamma$ by the update and conjunction axiom. By item 4 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent to $(M^c,\Gamma)\models([\mathsf{U},\mathsf{e}]\psi\wedge[\mathsf{U},\mathsf{e}]\chi)$. By the semantics this is equivalent to $(M^c,\Gamma)\models[\mathsf{U},\mathsf{e}](\psi\wedge\chi)$.

the case for $[\mathsf{U},\mathsf{e}][a]\psi$**:** Suppose $[\mathsf{U},\mathsf{e}][a]\psi \in \Gamma$. Given that $[\mathsf{U},\mathsf{e}][a]\psi \in \Phi$, $[\mathsf{U},\mathsf{e}][a]\psi \in \Gamma$ is equivalent to $(\mathsf{pre}(\mathsf{e}) \to [a][\mathsf{U},\mathsf{e}]\psi) \in \Gamma$ by the update and knowledge axiom. By item 5 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent to $(M^c,\Gamma) \models \mathsf{pre}(\mathsf{e}) \to [a][\mathsf{U},\mathsf{e}]\psi$. By the semantics this is equivalent to $(M^c,\Gamma) \models [\mathsf{U},\mathsf{e}][a]\psi$.

the case for $[\mathsf{U},\mathsf{e}][B^+]\psi$**:** Suppose $[\mathsf{U},\mathsf{e}][B^+]\psi \in \Gamma$. Given that $[\mathsf{U},\mathsf{e}][B^+]\psi \in \Phi$, $[\mathsf{U},\mathsf{e}][B^+]\psi \in \Gamma$ iff for all $\mathsf{f} \in \mathsf{U}$ every $B\mathsf{Uef}$-path from Γ ends in a $[\mathsf{U},\mathsf{f}]\psi$-state by item 7 of Lemma 25. By item 6 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent saying that for all $\mathsf{f} \in \mathsf{U}$ every $B\mathsf{Uef}$-path from Γ ends in a state where $[\mathsf{U},\mathsf{f}]\psi$ is true. By the semantics, this is equivalent to $(M^c,\Gamma) \models [\mathsf{U},\mathsf{e}][B^+]\psi$.

the case for $[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\psi$**:** Suppose $[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\psi \in \Gamma$. Given that $[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\psi \in \Phi$, $[\mathsf{U},\mathsf{e}][\mathsf{U}',\mathsf{e}']\psi \in \Gamma$ is equivalent to $[(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')]\psi \in \Gamma$ by the update composition axiom. By item 7 of Lemma 27 we can apply the induction hypothesis. Therefore this is equivalent to $(M^c,\Gamma) \models [(\mathsf{U},\mathsf{e});(\mathsf{U}',\mathsf{e}')]\psi$. This is equivalent to $(M^c,\Gamma) \models [\mathsf{U},\mathsf{e}][\mathsf{U},\mathsf{e}']\psi$.

Now it is easy to prove the completeness theorem.

Theorem 29 (completeness) For every $\varphi \in \mathscr{L}$

$$\models \varphi \text{ implies } \vdash \varphi$$

The proof is a slight variation on the proof presented in [4] (also see [27] for a textbook version).

Proof By contraposition. Suppose $\nvdash \varphi$. Therefore $\{\neg\varphi\}$ is a consistent set. By the Lindenbaum Lemma there is a set Γ which is maximal consistent in $cl(\neg\varphi)$ such that $\neg\varphi \in \Gamma$. By the Truth Lemma $(M^c,\Gamma) \models \neg\varphi$. Therefore $\not\models \varphi$.