ELSEVIER

Artificial Intelligence 82 (1996) 75-127

Artificial Intelligence

ALX, an action logic for agents with bounded rationality

Zhisheng Huang [1], Michael Masuch [2], László Pólos

Center for Computer Science in Organization and Management (CCSOM), University of Amsterdam, Sarphatistraat 143, 1018 GD Amsterdam, Netherlands

Received June 1993; revised November 1994

Abstract

We propose a modal action logic that combines ideas from H.A. Simon's *bounded rationality*, S. Kripke's *possible world semantics*, G.H. von Wright's *preference logic*, Pratt's *dynamic logic*, Stalnaker's *minimal change* and more recent approaches to *update semantics*. ALX (the xth action logic) is sound, complete and decidable, making it the first complete logic for two-place preference operators. ALX avoids important drawbacks of other action logics, especially the counterintuitive necessitation rule for goals (every theorem must be a goal) and the equally counterintuitive closure of goals under logical implication.

1. Introduction

Action logics are usually developed for (hypothetical) use by intelligent robots [6, 14, 40, 59, 71], as a description language of program behavior [24, 47], or as a contribution to philosophical logic [64]. Our effort is motivated by a different concern. We want to develop a formal language for social science theories, especially for theories of organizations. The difference in motivation leads to a new approach to action logic. We combine ideas from various strands of thought, notably H.A. Simon's notion of *bounded rationality*, G.H. Wright's approach to *preferences*, Kripke's *possible world semantics* in combination with binary modal operators, Pratt's *dynamic logic*, Stalnaker's notion of *minimal change*, and more recent ideas from *belief revision* and *update semantics* [15, 25]. Although fairly simple in its construction, ALX is good at handling some crucial problems of action logic. In particular, ALX avoids the counterintuitive necessitation

[1] E-mail: huang@ccsom.uva.nl.

[2] Corresponding author. Fax: (31-20) 5252800. E-mail: michael@ccsom.uva.nl.

SSDI 0004-3702(94)00090-5

rule for goals (every theorem must be a goal) and the equally counterintuitive closure of goals under logical implication. ALX is complete, decidable and enjoys the finite model property.

2. The framework for ALX

Most social science theories are expressed in natural language. They lack a formal scaffold that would allow one to check their consistency in a rigorous fashion, or to disambiguate natural language statements. As a consequence, these theories have acquired a reputation for "softness"—a soft way of saying that their logical properties are somewhat dubious. Reformulating them in a formal language with known properties would make consistency checking or disambiguation easier. Also, it would pave the way for other tasks, such as the examination of a theory's deductive closure properties. Understanding the deductive closure properties of a set of formulas is essential to automating the generation of theories from a given set of assumptions and introducing AI into theory building [45].

We focus on action logic as a formal language, because actions of individual or collective agents are key to the understanding of social phenomena. In fact, most social scientists agree that adequate theories of social relations must be action theories [1, 12, 16, 39, 52, 61]. Yet actions engender change and change is notoriously hard to grasp in the extensional context of first order languages [11]. This drives our attempt to develop a new logic, rather than taking First-Order Logic off the shelf. We call the new logic ALX (the xth action logic).

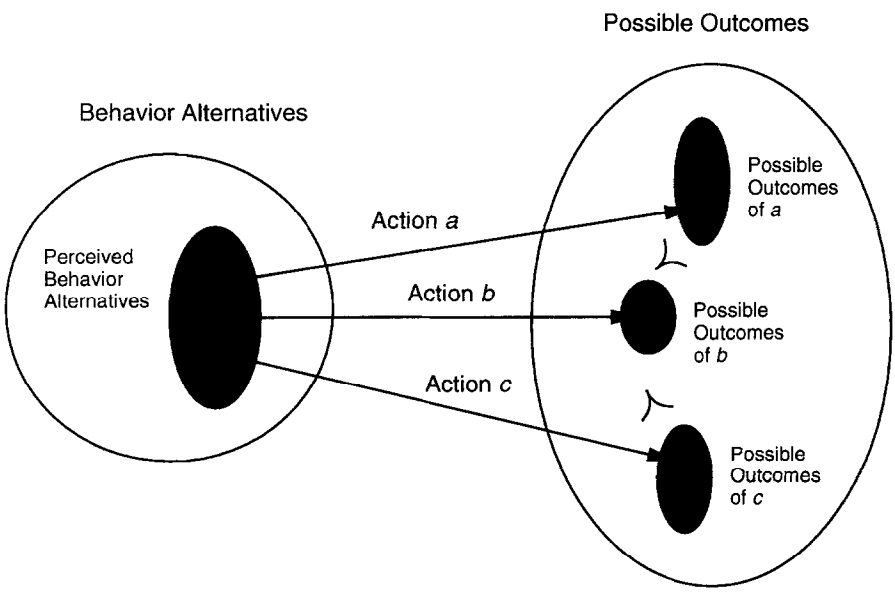

Fig. 1. Simon's bounded rationality.

Herbert A. Simon's conceptualization of *bounded rationality* [65] serves as a point of departure (see Fig. 1). His approach is intuitively appealing and had great impact on the postwar social sciences. Simon wanted to overcome the omniscience claims of the traditional conceptualizations of rational action. He assumed (1) an agent, with (2) a set of behavior alternatives, (3) a set of future states of affairs (each such state being the outcome of a choice among the behavior alternatives), and (4) a preference order over future states of affairs. The omniscient agent, endowed with "perfect rationality", would know all behavior alternatives and the exact outcome of each alternative; the agent would also have a complete preference order for those outcomes. An agent with bounded rationality, in contrast, would not know all the alternatives, would not know the exact outcome of each, and would not have a complete preference order for those outcomes.

Kripke's *possible world semantics* provides a natural setting for Simon's conceptualization. We assume a set of possible worlds with various relations defined over this set (we may also call those possible worlds *states*). One can see a behavior alternative as a mapping from states to states, so each behavior alternative constitutes an accessibility relation. An accessibility relation, in turn, can be interpreted as an opportunity for action, i.e., as an opportunity for changing the world by moving from a given state to another state. Accessibility relations are expressed by one-place modal operators, as in dynamic logic [24]. For example, the formula $\langle a \rangle \phi$ would express the fact that the agent has an action a at his disposal such that effecting a in the present situation would result in the situation denoted by proposition ϕ.

The perfectly rational agent would have a complete description of his actual state, a complete knowledge of all accessibility relations and a complete preference order over states. Agents with bounded rationality are less well informed. They may have an incomplete description of their actual state (we call those descriptions *situations*), incomplete knowledge of the accessibility relations, and an incomplete preference order over situations. In particular, one can assume that their knowledge and the range of their actions is finite. As it turns out, this assumption is crucial for defining a straightforward concept of goal in ALX's object language.

Situations are represented as sets of states and expressed by propositions. Propositions, in turn, denote the set of states where they obtain. So, the more specific an agent's knowledge about a situation, the more detailed the propositional description of that situation would be. The limit case, a complete description, would uniquely identify one state. Less specific descriptions would lack that uniqueness, identifying the set of those states where the description would hold (but remaining uncommitted about other "aspects" not covered by the description). This is the standard approach to representing incomplete information, used in denotational semantics [62,63] and epistemic logic [20].[3]

[3] Framing bounded rationality in terms of possible worlds semantics reveals a fine point usually ignored: one can see that omniscience—the limiting case—is contingent upon the choice of the language. Full rationality in an absolute sense would require a language isormorphic to the universe "out there", but such a language is not available. Any formal theory about full rationality has to make simplifying assumptions about the world, but those assumption, by their simplifying nature, violate the ontology of full rationality in some sense or another.

Preferences—not goals—provide the basis for rational action in ALX. Following von Wright [72], a preference statement is understood as a statement about situations. For example, the statements that "I prefer oranges to apples" is interpreted as the fact that "I prefer the states in which I have an orange to the states in which I have an apple." Following von Wright again, we assume that an agent who claims to prefer oranges to apples should prefer a situation where he has an orange but *no* apple to a situation where he has an apple but *no* orange. Preferences are expressed via two-place modal operators; if the agent prefers the proposition ϕ to the proposition ψ, we write $\phi \boldsymbol{P} \psi$.

Normally, the meaning of a preference statement is context-dependent, even when this is not made explicit. An agent may claim to prefer an apple to an orange—and actually mean it—but he may prefer an orange to an apple later—perhaps because then he already had an apple. To capture this context dependency, we borrow the notion of minimal change from Stalnaker's approach to conditionals [68]. The idea is to apply the conjunction expansion principle only to situations that are minimally different from the agent's present situation—just as different as they really need to be in order to make the propositions true about which preferences are expressed. We introduce a binary function, *cw*, to the semantics that determines a set of "closest" states relative to a given state, such that the new states fulfill some specified conditions, but resemble the old state as much as possible in all other respects. For situations (sets of states), we apply *cw* to each element of the situation separately. This allows us to avoid some technical problems arising in conditional logic [38, 50, 68].[4]

ALX provides a complete syntactic characterization of preferences, so one can derive new preference statements from old ones by using its machinery. Closing the set of preference statements under the rules of inferences yields a preference order that serves as the basis for deriving goals. *Goals*, in turn, are defined in terms of preferences and accessibilities (we will argue that there are several plausible goal definitions, corresponding to increasingly stronger notions of rationality). Note that goals need not be unique; this follows from the fact that world descriptions and preference orders are usually incomplete. Also, the goal set need not be closed under logical implication, so agents need not treat undesired consequences of desired outcomes (e.g., pain as a consequence of having one's teeth repaired) as goals (as opposed to action logics that use the concept of "goal" as the primitive notion of rational guidance).

3. Formal syntax and semantics

3.1. Syntax

ALX is a multimodal propositional logic. The propositional alphabet consists of a countable set of lower-case Latin symbols p_i to denote primitive propositions. The action alphabet has a finite set of actions a_i. Lower-case Greek letters $\phi, \psi, \rho, \ldots$ (with or without subscript) denote well-formed formulae. We have $\langle a \rangle \psi$ to denote the one-place existential accessibility relation for action a and $\boldsymbol{P}$ to denote the two-place preference

[4] The reader is referred to the discussion part of this paper for details.

relation. $\circ$ serves as two-place operator for *updates*; updates are changes caused by an action. Note that updates in ALX refer to real state changes, not epistemological ones [15], so an update does not produce a new knowledge state, but a new situation.

Definition 1 (*Syntax*). Let $ATOM = \{p_i : i < \omega\}$ and $ACTION = \{a_1, \ldots, a_k\}$ for some $k \in \omega$ with ω standing for the ordinality of natural numbers. The set of formulae FML is defined recursively as follows:

- $ATOM \subseteq FML$,
- $\phi \in FML \Rightarrow \neg\phi \in FML$,
- $\phi, \psi \in FML \Rightarrow (\phi \wedge \psi) \in FML$,
- $\phi \in FML, a \in ACTION \Rightarrow (\langle a \rangle \phi) \in FML$,
- $\phi, \psi \in FML \Rightarrow (\phi \circ \psi) \in FML$,
- $\phi, \psi \in FML \Rightarrow (\phi \boldsymbol{P} \psi) \in FML$.

Define $\bot$ as $\phi \wedge \neg\phi$ for an arbitrary ϕ, and $[a]\phi$ as $\neg\langle a \rangle \neg\phi$. Define the Boolean connectives $\{\vee, \rightarrow, \leftrightarrow\}$ and the truth constant $\top$ from the given Boolean connectives in the usual way.

3.2. Semantics

Definition 2 (*ALX models*). We call $M = \langle W, cw, \succ, \{R^a\}_{a \in ACTION}, V \rangle$ an ALX model if:

- W is a set of possible worlds,
- $cw : W \times \mathcal{P}(W) \rightarrow \mathcal{P}(W)$ is a closest world function,
- $\succ \subseteq \mathcal{P}(W) \times \mathcal{P}(W)$ is a comparison relation for preferences,
- $R^a \subseteq W \times W$ is an accessibility relation for each a in $ACTION$,
- $V : ATOM \rightarrow \mathcal{P}(W)$ is an assignment function for primitive propositions,

and if M satisfies the following conditions:

(CS1) $cw(w, X) \subseteq X$.

(CS2) $w \in X \Rightarrow cw(w, X) = \{w\}$.

(CSC) $cw(w, X) \cap Y \subseteq cw(w, X \cap Y)$.

(NORM) $(\emptyset \not\succ X)$, $(X \not\succ \emptyset)$.

(TRAN) $cw(w, X \cap \overline{Y}) \succ cw(w, Y \cap \overline{X})$ and $cw(w, Y \cap \overline{Z}) \succ cw(w, Z \cap \overline{Y})$
$\Rightarrow cw(w, X \cap \overline{Z}) \succ cw(w, Z \cap \overline{X})$,
where $\overline{Y} = W - Y$.

(CS1), (CS2) and (CSC) constrain the closest world function. (CS1) ensures that the closest ϕ-worlds (relative to a given world) are indeed ϕ-worlds; (CS2) ensures that w is its own (and unique) closest ϕ-world if ϕ is true at w. (CSC) says that if ψ is true at the closest ϕ-world, then the closest ϕ-world is also a closest ϕ-and-ψ-world. (NORM) and (TRAN) constrain the semantic preference relation. They require normality and transitivity; "normality" stipulates that no comparison between two sets of worlds would involve an empty set of worlds.

In the following, we will use $M = \langle W, cw, \succ, R^a, V\rangle$ to denote $M = \langle W, cw, \succ, \{R^a\}_{a\in ACTION}, V\rangle$ if the omission causes no ambiguity.

Definition 3 (*Interpretation function*). Let *FML* be as above and let $M = \langle W, cw, \succ, R^a, V\rangle$ be an ALX model. The interpretation function $[\![\ ]\!]_M : FML \to \mathcal{P}(W)$ is defined as follows:

$$
\begin{aligned}
[\![p_i]\!]_M &= V(p_i),\\
[\![\neg\phi]\!]_M &= W \setminus [\![\phi]\!]_M,\\
[\![\phi \wedge \psi]\!]_M &= [\![\phi]\!]_M \cap [\![\psi]\!]_M,\\
[\![\langle a\rangle\phi]\!]_M &= \{w \in W : \exists w' \in W(R^a w w' \text{ and } w' \in [\![\phi]\!]_M)\},\\
[\![\phi \circ \psi]\!]_M &= \{w \in W : \exists w' \in W(w' \in [\![\phi]\!]_M \text{ and } w \in cw(w', [\![\psi]\!]_M))\},\\
[\![\phi \boldsymbol{P} \psi]\!]_M &= \{w \in W : cw(w, [\![\phi \wedge \neg\psi]\!]_M) \succ cw(w, [\![\neg\phi \wedge \psi]\!]_M)\}.
\end{aligned}
$$

The interpretation of the primitive propositions and the Boolean connectives is straightforward. A proposition letter p_i evaluates to the set of worlds where p_i obtains, the negation of a proposition ϕ evaluates to the complement of the ϕ-worlds, and the conjunction of ϕ and ψ evaluates to the intersection of the ϕ-worlds and the ψ-worlds. The interpretation of $\langle a\rangle\phi$ yields the set of worlds from where the agent can access at least one ϕ-world via action a. We use the "existential" version of the action modality, because real-life decisions typically depend on the *possibility* of a specific action in a specific situation.

The interpretation of $\phi \circ \psi$ yields the set of worlds where ψ holds so that one could have got there from a closest ϕ-world. Note that $\phi \circ \psi$ is a backward-looking operator [15]. Note also that $\phi \circ \psi$ is closely related to the intensional conditional ("wiggle") known from Stalnaker's and D. Lewis' work [38,68], where $\phi \rightsquigarrow \psi$ is meant to express "if ϕ were the case, then ψ would be the case".[5] One can express (although not term-define) the wiggle in terms of the update operator via the so-called *Ramsey rule* [15]:

$$\vdash (\chi \to (\phi \rightsquigarrow \psi)) \Leftrightarrow \vdash ((\chi \circ \phi) \to \psi).$$

The interpretation of $\phi \boldsymbol{P} \psi$ yields the set of worlds where the agent prefers (at each of those worlds) the closest ϕ-and-not-ψ-worlds to the closest ψ-and-not-ϕ-worlds.

Define the forcing relation as:

$$M, w \Vdash \phi \overset{\text{def}}{\Longleftrightarrow} w \in [\![\phi]\!]_M.$$

Definition 4 (*The logic ALX*). Let *FML* be as above, let *Mod* be the class of all ALX models and let $[\![\ ]\!]_M$ be as above too, defined for every model $M \in Mod$. We call the logic ALX = $\langle FML, Mod, [\![\ ]\!]_M\rangle$ ALX logic.

[5] In fact, in Stalnaker, the "wiggle" is a "corner" ($>$); we prefer the $\rightsquigarrow$ because it frees the $>$ for other uses.

Define the semantic consequence relation, $\models$, as usual:

$$M = \langle W, cw, \succ, R^a, V\rangle \models \phi \overset{\text{def}}{\Longleftrightarrow} (\forall w \in W)(M, w \Vdash \phi).$$

$$M \models \Gamma \overset{\text{def}}{\Longleftrightarrow} (\forall \gamma \in \Gamma)(M \models \gamma).$$

$$Mod(\Gamma) \overset{\text{def}}{\Longleftrightarrow} \{M \in Mod : M \models \Gamma\}.$$

$$\Gamma \models \phi \overset{\text{def}}{\Longleftrightarrow} Mod(\Gamma) \subseteq Mod(\{\phi\}).$$

Definitions 2–4 provide a semantic characterization of ALX. The next definition provides a complete syntactic characterization.

Definition 5 (*ALX inference system*). Let ALXS be the following set of axioms and rules of inference.

(BA)	*all propositional tautologies.*	
(A1)	$\langle a\rangle\bot$	$\leftrightarrow \bot$
(A2)	$\langle a\rangle(\phi \vee \psi)$	$\leftrightarrow \langle a\rangle\phi \vee \langle a\rangle\psi$
(A3)	$\langle a\rangle(\phi \wedge \psi)$	$\rightarrow \langle a\rangle\phi \wedge \langle a\rangle\psi$
(U1)	$\phi \circ \psi$	$\rightarrow \psi$
(U2)	$\phi \wedge \psi$	$\rightarrow \phi \circ \psi$
(U3)	$\neg(\phi \circ \bot), \neg(\bot \circ \phi)$	
(U4)	$(\phi \vee \psi) \circ \chi$	$\leftrightarrow \phi \circ \chi \vee \psi \circ \chi$
(U5)	$(\phi \wedge \psi) \circ \psi$	$\rightarrow \phi$
(U6)	$(\phi \circ \psi) \wedge \chi$	$\rightarrow \phi \circ (\psi \wedge \chi)$
(CEP)	$\phi \boldsymbol{P} \psi$	$\leftrightarrow (\phi \wedge \neg\psi)\boldsymbol{P}(\neg\phi \wedge \psi)$
(TR)	$(\phi \boldsymbol{P} \psi) \wedge (\psi \boldsymbol{P} \chi)$	$\rightarrow (\phi \boldsymbol{P} \chi)$
(N)	$\neg(\bot \boldsymbol{P} \phi), \neg(\phi \boldsymbol{P} \bot)$	
(MP)	$\vdash \phi$ & $\vdash \phi \rightarrow \psi$	$\Rightarrow \vdash \psi$
(NECA)	$\vdash \phi$	$\Rightarrow \vdash [a]\phi$
(MONA)	$\vdash \langle a\rangle\phi$ & $\vdash \phi \rightarrow \psi$	$\Rightarrow \vdash \langle a\rangle\psi$
(MONU)	$\vdash \phi \circ \psi$ & $\vdash \phi \rightarrow \phi'$	$\Rightarrow \vdash \phi' \circ \psi$
(SUBA)	$\vdash (\phi \leftrightarrow \phi')$	$\Rightarrow \vdash (\langle a\rangle\phi) \leftrightarrow (\langle a\rangle\phi')$
(SUBU)	$\vdash (\phi \leftrightarrow \phi')$ & $\vdash (\psi \leftrightarrow \psi')$	$\Rightarrow \vdash (\phi \circ \psi) \leftrightarrow (\phi' \circ \psi')$
(SUBP)	$\vdash (\phi \leftrightarrow \phi')$ & $\vdash (\psi \leftrightarrow \psi')$	$\Rightarrow \vdash (\phi \boldsymbol{P} \psi) \leftrightarrow (\phi' \boldsymbol{P} \psi')$

Most axioms are straightforward. As usual, we have the propositional tautologies (BA). Since ALX is a *normal* modal logic, the absurdum is not true anywhere, so it is not accessible (A1). The action modalities behave as usual, so they distribute over disjunction both ways, but over conjunction only in one direction (A2) and (A3).

Indeed, we can get to ϕ-or-ψ-worlds via action a if and only if we can get via a to a ϕ-world or to a ψ-world. However, being able to get to ϕ-worlds via action a *and* being able to get to ψ-worlds via action a does not necessarily mean that a can get us to a world that is both ϕ and ψ.

As mentioned above, $\circ$ is a backward-looking operator. So, a successful ψ-update ends up in a ψ-world (U1) and the truth of both ϕ and ψ at a world allows us to perform a vacuous ψ-update, i.e., to remain at that world (U2). (U3) reiterates the normality condition for updates. Since there is no world where the absurdum is true, an update with the absurdum cannot succeed. (U4) posits the left distribution of the disjunction over the update operator. The intuition is that if we have got to a χ-world from a ϕ- *or* a ψ-world, we have updated either from a ϕ-world or from a ψ-world. (U5) tells us that a void update is not going to change any conditions. (U6) posits that if χ holds after updating ϕ with ψ, then we can update ϕ with $\psi \wedge \chi$ and obtain the same result. Readers more familiar with closest world functions may already sense how the update operator will mimic the closest world function in the syntax, helping to construct a canonical model in the completeness proof.

The axioms for the preference operator posit the conjunction expansion principle (CEP), transitivity (TR), and normality (N). So, if we prefer ϕ to ψ, we will also prefer the absence of ψ to the absence of ϕ. If we prefer ϕ to ψ, we are apt to prefer ϕ-and-not-ψ to ψ-and-not-ϕ. We have transitivity because we think that it is a natural principle of preference orders. Normality is required to avoid inconsistent preference statements. For example, without normality, we get a violation of irreflexivity via the only-if part of the contraposition principle.[6] We need not state irreflexivity as an axiom, since it is derivable from (CEP) and (N). By the same token, we can derive contraposition and asymmetry for the preference operator.

Proposition 6 (More properties of the preference operator). *The following formulas are theorems of ALX*:

(CP) $\phi \boldsymbol{P} \psi \leftrightarrow (\neg\psi)\boldsymbol{P}(\neg\phi)$,

(IR) $\neg(\phi \boldsymbol{P} \phi)$,

(NT) $\neg(\top \boldsymbol{P} \phi), \neg(\phi \boldsymbol{P} \top)$,

(AS) $\phi \boldsymbol{P} \psi \rightarrow \neg(\psi \boldsymbol{P} \phi)$.

Proof. See Appendix A. □

We have modus ponens and the necessitation rule for the universal action modality (NECA) and monotonicity for the existential action modality. For the update operator, we have *left* monotonicity, but not *right* monotonicity, the intuition being that a move from a ϕ-world to the closest ψ-world w might end up at a different world than the move to the closest ψ'-world even if ψ implies ψ' at w. Logically equivalent propositions are

[6] An alternative method for preserving consistency, suggested by von Wright [72] and used by Hansson [22], requires the "independence" of propositions in certain axioms. We have used this approach in [28], but it is less straightforward because the definition of "independence" is nontrivial.

substitutible in action, update and preference formulas (SUBA), (SUBU), (SUBP). Note that we do *not* have monotonicity for preferences. Because of this, we are able to avoid the counterintuitive deductive closure of goals.

4. Formal properties of ALX

ALX has pleasant logical properties. We have:

Proposition 7 (Soundness of ALXS). *ALXS is sound, i.e., for an arbitrary set of formulas Δ and an arbitrary formula ϕ,*

$$\Delta \vdash_{\mathrm{ALX}} \phi \Rightarrow \Delta \models \phi.$$

Proof. See Appendix A. □

Next, we have:

Proposition 8 (Completeness of ALXS). *ALX is complete, i.e., for an arbitrary set of formulas Δ and an arbitrary formula ϕ,*

$$\Delta \models \phi \Rightarrow \Delta \vdash_{\mathrm{ALX}} \phi.$$

Proof. See Appendix A. □

Furthermore, ALX is decidable. Stronger even, we have the finite model property for ALX, that is, for each non-theorem ψ there exists a finite model that provides a counterexample for ψ. Since ALX is recursively axiomatizable, this means that ALX is decidable.

Definition 9 (*Finite model property*). A logic S is said to have the finite model property, iff, for arbitrary ϕ such that $\nvdash_S \phi$, there exists a finite model M such that:

(1) $\exists w(M, w \Vdash \neg\phi)$,

(2) $\forall\rho(\vdash_S \rho \Rightarrow \forall w(M, w \Vdash \rho))$.

Theorem 10. *ALX has the finite model property.*

Proof. See Appendix A. □

ALXS is finite, so ALX is finitely axiomatizable. The finite model property together with finite axiomatizability imply decidability (cf. [31, p. 153]. As a consequence, we have:

Corollary 11. *ALX is decidable.*

5. Applying ALX

ALX provides considerable flexibility in defining new modal operators by using the three primitive operators. We concentrate on operators of potential use in defining goals.

In the following, we assume that the preference order of an agent is finite and hence the corresponding set of preference statements. Call this set Σ_P. Recall furthermore that the range of action alternatives is finite, too (as stipulated in Definition 1). Suppose that $\Sigma_P = \{\psi_1, \ldots, \psi_n\}$. In the following, we use the notation $\forall\psi\Phi(\psi)$ (where $\Phi(\psi)$ is an formula that contains ψ) to denote $\Phi(\psi_1) \wedge \cdots \wedge \Phi(\psi_n)$ and $\exists\psi\Phi(\psi)$ to denote $\Phi(\psi_1) \vee \cdots \vee \Phi(\psi_n)$.

Define accessibility as follows:

Definition 12 (*Accessibility*). Let $\boldsymbol{A}\phi$ stand for the fact that situation ϕ is accessible via an action. Define:

$$\boldsymbol{A}\phi \stackrel{\text{def}}{\Longleftrightarrow} \langle a_1\rangle\phi \vee \langle a_2\rangle\phi \vee \cdots \vee \langle a_k\rangle\phi.$$

Thus, operator $\boldsymbol{A}$ acts as an existential quantifier over action terms; if situation ϕ is accessible via an arbitrary action, then we have $\boldsymbol{A}\phi$. Note how bounded rationality impinges on this definition. In defining accessibility we can stay inside the object language because we assume that agents are not omnipotent and have only finitely many action alternatives at their disposal.

Define a "good" situation ϕ as a situation that the agent prefers to its negation and conversely for a "bad" situation.

Definition 13 (*Good, bad situations*). Let $\boldsymbol{GO}\phi$ stand for a "good" situation ϕ and $\boldsymbol{BA}\phi$ for a "bad" situation ϕ. Define:

$$\boldsymbol{GO}\phi \stackrel{\text{def}}{\Longleftrightarrow} \phi\boldsymbol{P}\neg\phi, \qquad \boldsymbol{BA}\phi \stackrel{\text{def}}{\Longleftrightarrow} \neg\phi\boldsymbol{P}\phi.$$

Define an element of the agent's preference order in the obvious way:

Definition 14 (*Element of the preference order*). Let $\boldsymbol{PO}\phi$ stand for an element in the agent's preference order. Define:

$$\boldsymbol{PO}\phi \stackrel{\text{def}}{\Longleftrightarrow} (\phi\boldsymbol{P}\rho) \vee (\rho\boldsymbol{P}\phi).$$

5.1. Goals

Goals are a crucial notion for action logics. Following the basic notions of bounded rationality (and, for that matter, standard decision theory), we derive goals from preferences; they are not a primitive notion as in other action logics. But there are many ways to base goals on preferences. A situation may be singled out as a goal simply because it is better than its negation, or, perhaps, because it is better than other situations; it may be satisficing, outstanding (extremal), or optimal. Bounded rationality is often identified

with the notion that agents do not optimize, at least not in the sense of putting much energy into the search for extremal solutions; instead, they are said to satisfice. However, the reduction of bounded rationality to satisficing is misleading. Satisficing is, indeed, relevant when the existence, or the accessibility, of potential goal states is *unknown*. If a known alternative meets a given aspiration level, then, as a rule, the agent will not *search* for a better state; conversely, if no known alternative meets the aspiration level, the agent will search for better solutions, at least up to a certain point. However, agents might act irrational if they do not pursue *known* better accessible alternatives; if they never do, aspiration levels could only go down). Bounded rationality has been introduced in order to develop a more realistic framework of rational decision making, and a drive for *improving* one's situation is apparent in many human decisions.

We present four goal definitions (good, satisficing, extremal, optimal), and discuss some obvious modifications of these definitions.

Agents might opt for a state simply because it is better than its negation, particularly if only a few alternatives are considered. For example, if an agent finds himself late at night far from home without a car, he might base his decision to take a taxi on the simple deliberation that it is better to take a taxi than not to take a taxi. A "good" goal can be defined by using the "good" operator $\boldsymbol{GO}$:

Definition 15 (*Good goal*). Let $\boldsymbol{G}^{\mathrm{g}}\phi$ denote the fact that ϕ is a good goal. Define:

$$\boldsymbol{G}^{\mathrm{g}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} \boldsymbol{GO}\phi.$$

Thus a good goal is a situation that is preferred to its negation.

The second definition involves a satisficing goal. As noted above, satisficing—important as it is—is a procedural addendum to the definition of bounded rationality. Whereas the declarative part of bounded rationality concerns incomplete knowledge, the procedural part concerns the question of what to do when the knowledge is not complete enough [67]. So, satisficing states are, in fact, satisfactory states made accessible via search. Let $\boldsymbol{S}\phi$ stand for an arbitrary satisficing situation ϕ and relax the definition of action terms by allowing for mnemonic expressions:

$$\boldsymbol{S}\phi \Leftrightarrow \langle \textit{satisficing-search} \rangle \phi \wedge \boldsymbol{PO}\phi.$$

Note that this definition does not exclude the possibility that the search is void in cases that the satisficing solution is already at hand. Note also that the definition is using the name of search, but is not defining, or describing, the search process itself. Define a satisficing goal in terms of a satisficing state:

Definition 16 (*Satisficing goal*). Let $\boldsymbol{G}^{\mathrm{s}}\phi$ denote the fact that ϕ is a satisficing goal. Define:

$$\boldsymbol{G}^{\mathrm{s}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} \boldsymbol{S}\phi.$$

As argued above, agents may try to maximize, or even optimize, if the context supports the search for extremal values. For example, in production planning, optimal

solutions are sought and implemented on a daily basis. Whether a solution is maximal or optimal depends, of course, on the structure of the preference order of an agent. If it is partial, but not total, the order may contain several maximal, incomparable elements. If, furthermore, more than one maximal element is accessible, then an optimal goal (in the intuitive sense of a best overall solution) cannot be defined. Conversely, an optimal goal can be identified if the order is total and at least one situation is accessible. By the same token, a partial order gives rise to an optimal goal if only one maximal situation is accessible. We define a "best choice" as a maximal goal and specify the conditions under which such a best choice may, in fact, be optimal.

Definition 17 (*Maximal goal*). Let $G^{\mathrm{bc}}\phi$ denote the fact that ϕ is a best choice. Define:

$$G^{\mathrm{bc}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} PO\phi \wedge \forall\chi(\chi P\phi \rightarrow \neg A\chi).$$

Thus, a best choice is an accessible situation to which no other accessible situation is preferred. A best choice ϕ is optimal, if ϕ is unique:

Definition 18 (*Optimal goal*). Let $G^{\mathrm{op}}\phi$ denote the fact that ϕ is optimal. Define:

$$G^{\mathrm{op}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} G^{\mathrm{bc}}\phi \wedge \forall\psi((G^{\mathrm{bc}}\psi) \rightarrow (\phi \leftrightarrow \psi)).$$

Ironically, best choices need not be good nor satisficing. In a tight spot, an agent's best alternative might simply be the best among dubious alternatives.

The above definitions can be modified according to the domain. For example, a stronger notion of rationality may require that goals be consistent, so that agents will not select both ϕ and $\neg\phi$ as goals in the same situation. We did not require consistency upfront, because there are many applications of bounded rationality that do allow for contradictory goals [41,42], but consistency can be built into the goal definitions by requiring that a goal be, at least, a good goal. Because of the irreflexivity of the P operator, good goals are always contradiction-free. Define a consistent satisficing goal as follows:

Definition 19 (*Satisficing Consistent Goal*). Let $G^{\mathrm{sc}}\phi$ denote the fact that ϕ is a satisficing, consistent goal. Define:

$$G^{\mathrm{sc}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} GO\phi \wedge S\phi.$$

A consistent best choice can be defined analogously. Optimal goals are always consistent because they are unique.

Another reasonable modification of the goal definitions is obtained by imposing the requirement of accessibility. Sometimes, goals are pursued even when the agent is uncertain whether they are accessible (setting seemingly, but not really, inaccessible goals is sometimes hailed as post-modern management style [9]). In better-understood circumstances, however, accessibility may appear as a reasonable requirement for a

goal definition. All goal definitions above can be strengthened accordingly by adding the accessibility requirement. Another useful modification is obtained by distinguishing between *maintenance* and *achievement* goals. Again, it is easy to see how to do this: add the goal situation as a conjunct to the definiens of the respective goal definition in case of a maintenance goal and add the negation in case of an achievement goal. For example, a "good" achievement goal can be defined as follows:

Definition 20 (*Good achievement goal*). Let $\boldsymbol{G}^{\mathrm{ga}}\phi$ denote the fact that ϕ is a good achievement goal. Define:

$$\boldsymbol{G}^{\mathrm{ga}}\phi \stackrel{\mathrm{def}}{\Longleftrightarrow} \boldsymbol{G}^{\mathrm{g}}\phi \wedge \neg\phi.$$

It might go without saying that the definition of extremal goals must always make a stipulation about accessibility; otherwise, the sky is the only limit.

5.2. *Using goal definitions: an example*

Although the underlying propositional language imposes obvious limitations on the present version of the logic, ALX can already serve as a knowledge-representation tool. We demonstrate this by representing Max Weber's typology of rationality which informed large parts of twentieth-century sociology [16, 52, 53]. Max Weber distinguishes between three "types" of rationality: (1) *traditional* rationality, (2) *value* rationality, (3) *goal* rationality. Traditional rationality is circumscribed as a mode of behavior along stimulus response patterns. Agents follow the tradition, rather than seeking to improve their lives. So, if a situation is a goal, then it remains a goal, regardless of the precondition. Let $\boldsymbol{G}$ stand for an arbitrary goal; then we can characterize traditional rationality by the formula

$$\boldsymbol{G}\phi \circ \chi \rightarrow \boldsymbol{G}\phi.$$

The second type of rationality, value rationality, is circumscribed as adherence to preset goals that are singled out for their intrinsic value and without regard for possible ramifications. We can express this by stipulating that all goals must be good goals and, furthermore, that the ramifications of such goals do not count and hence, should not appear in the preference order:

$$\boldsymbol{G}\phi \rightarrow \boldsymbol{G}^{\mathrm{g}}\phi \wedge \forall\psi((\phi \rightsquigarrow \psi) \wedge \neg(\phi \leftrightarrow \psi) \rightarrow \neg \boldsymbol{PO}\psi).$$

The third type of rationality is circumscribed as the unconditional search for optimal solutions (*goal rationality*). In this rationality mode all goals are, at least, best choices:

$$\boldsymbol{G}\phi \rightarrow \boldsymbol{G}^{\mathrm{bc}}\phi.$$

Weber cautions his readers repeatedly against taking his typology for a complete classification. Our formal representation immediately shows, that, in fact, the set of building blocks entering the characterizations can give rise to many alternative rationality types. Much of the confusion about Weber's typology would, in fact, go away, if one would realize this more clearly.

6. Discussion

ALX provides the skeleton of a preference-driven action logic, based on a propositional description language and three types of modal operators, a preference operator, an update operator and a set of action modalities.

6.1. Preferences

Although preference-based decision making has received some attention in the AI-literature [8, 70], ALX is actually the first preference-based action logic.

Modal preference logic was introduced by Halldén [17] and codified by von Wright [72, 73], whose preference operator satisfies irreflexivity, transitivity and the conjunction expansion principle. We modified his approach by adding normality and making preferences context-dependent. Our choice of *irreflexivity* was made for technical reasons. The machinery for reflexive preferences is more complex; also, reflexive preference statements are less intuitive (indeed, it is not easy to express reflexive preferences in natural language). *Transitivity* of preferences is widely seen as a basic requirement of rational, preference-based decision making and has not been challenged in the basic setup of bounded rationality. However, more radical applications of bounded rationality have done away with transitivity, claiming that organizational choice is often intransitive [42]. ALX can accommodate intransitive preferences since transitivity can be dropped without losing completeness and decidability [29]. This *non*-transitive logic does, in fact, allow for *in*transitive preferences, as the following example shows. Call an ALX logic without (TR) $\mathrm{ALX}^{-\mathrm{TR}}$.

Claim 21. *There exists an* $ALX^{-\mathrm{TR}}$ *model* $M = \langle W, cw, \succ, R^a, V\rangle$ *and a world* $w \in W$ *such that preferences are not transitive, even though the comparison relation* $\succ$ *is transitive. In particular, we claim that* $M, w \models (p\boldsymbol{P}q) \wedge (q\boldsymbol{P}r) \wedge \neg(p\boldsymbol{P}r)$.

Proof. Suppose that the set of primitive propositions is $\{p, q, r\}$. We define the model $M = \langle W, cw, \succ, R^a, V\rangle$ as follows:

$$W = \{w_{pqr}, w_{pq}, w_{pr}, w_{qr}, w_p, w_q, w_r, w_\emptyset\}.$$

We define cw (to the extent that we need it for the example).

$$cw(w_p, [\![p \wedge \neg q]\!]_M) = \{w_p\}.$$
$$cw(w_p, [\![\neg p \wedge q]\!]_M) = \{w_q\}.$$
$$cw(w_p, [\![q \wedge \neg r]\!]_M) = \{w_{pq}\}.$$
$$cw(w_p, [\![\neg q \wedge r]\!]_M) = \{w_{pr}\}.$$
$$cw(w_p, [\![p \wedge \neg r]\!]_M) = \{w_p\}.$$
$$cw(w_p, [\![\neg p \wedge r]\!]_M) = \{w_r\}.$$

$$\succ = \{\langle w_p, w_q\rangle, \langle w_{pq}, w_{pr}\rangle\}.$$

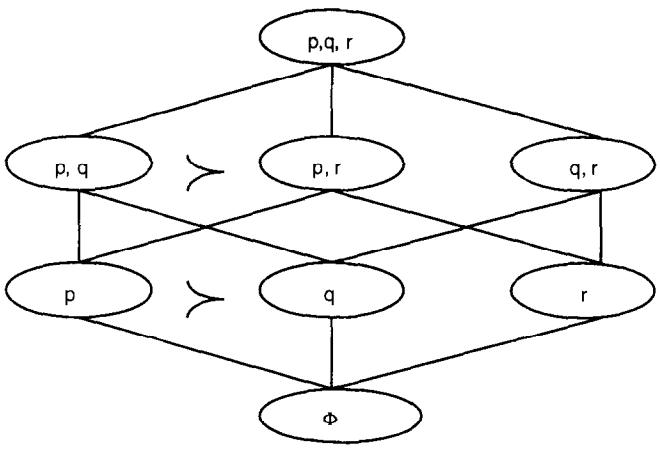

Fig. 2. Counterexample against the transitivity.

$$V(p) = \{w_{pqr}, w_{pq}, w_{pr}, w_p\}.$$
$$V(q) = \{w_{pqr}, w_{pq}, w_{qr}, w_q\}.$$
$$V(r) = \{w_{pqr}, w_{pr}, w_{qr}, w_r\}.$$

It is easy to see that the above model M is an $\mathrm{ALX}^{-\mathrm{TR}}$ model and that $M, w_p \models (p\boldsymbol{P}q) \wedge (q\boldsymbol{P}r) \wedge \neg(p\boldsymbol{P}r)$ (see also Fig. 2). □

Normality (N) was originally added to ALX in order to achieve completeness. Furthermore, (N) turns out to be instrumental in blocking counterexamples against the conjunction expansion principle.

The *conjunction expansion principle* (CEP) itself has always raised eyebrows and contributed to preference logic's unpopularity [49]. We have kept (CEP) for several reasons. First, we do not think that (CEP) restricts the logic's ability to represent preferences or preference-related notions, such as goals. At least, we cannot think of examples where it makes sense to prefer ϕ above ψ but not to prefer ϕ-and-not-ψ to ψ-and-not-ϕ. Second, we need (CEP) in the construction of context-dependent preferences. If the conjunction expansion principle is dropped from the semantics, the result is the following interpretation function of the preference operator:

$$M, w \Vdash \phi\boldsymbol{P}\psi \quad \text{iff} \quad cw(w, [\![\phi]\!]_M) \succ cw(w, [\![\psi]\!]_M),$$

and on this interpretation function, the following formula becomes valid:

$$\phi \wedge \psi \rightarrow \neg(\phi\boldsymbol{P}\psi).$$

But this formula is obviously counterintuitive. Third, ALX does not give rise to the traditional counterexamples against (CEP), because (N) blocks these examples. As we argue elsewhere [28], these examples are based on implicit partiality. For instance, assume, as in [4], that it is better that Smith *and* his wife are happy $(p \wedge q)$, than that Smith alone is happy: $(p \wedge q)\boldsymbol{P}p$. Conjunction expansion yields $(p \wedge q \wedge \neg p)\boldsymbol{P}\neg(p \wedge q) \wedge p$

and hence the preference for a contradictory state of affairs. This example comes out false in ALX because (N) assures that:

$$(\phi \wedge \psi)\boldsymbol{P}\phi \leftrightarrow \bot\boldsymbol{P}(\phi \wedge \neg\psi) \leftrightarrow \bot.$$

ALX forces the user to make the implication explicit that if Smith is happy alone, his wife is not happy: $(p \wedge q)\boldsymbol{P}(p \wedge \neg q)$. This statement entails no preference for a contradictory state of affairs; it is equivalent to its conjunction expansion. Other counterexamples to the conjunction expansion principle also exploit implicit partiality [4,21] and are blocked in the same way in ALX.[7]

ALX has a situational semantics for preference relations: the agent is supposed to hold a preference of ϕ above ψ iff she would prefer ϕ-and-not-ψ to ψ-and-not-ϕ under conditions as similar as possible to her actual situation. Obviously, situational preferences can be unstable; the agent may hold a specific preference in one situation and an opposite preference in another. Although unstable preferences play an important role in many applications of bounded rationality [3,41,42,51], stable preferences might still be handy for theoretical purposes (e.g., when using the logic to represent economic theories where stability of preferences is often assumed [10]). A stable preference relation would be one that does not change from world to world. We have discussed stable preferences elsewhere [28]; stability of the preference relation does obtain, for example, if a preference depends only on a finite (possibly empty) set of conditions that can be expressed as propositions. Stable preferences can be characterized with the following axiom:

$$\text{(UOP)} \quad (\phi\boldsymbol{P}\psi) \circ \chi \rightarrow (\phi\boldsymbol{P}\psi),$$

as shown in [28].

6.2. *Minimal change and actions*

Our notion of action is adopted from dynamic logic and its axioms (A1)–(A3) are not problematic. It should be clear, however, that this notion of action is, in a sense, contemplative: it answers the question of whether a particular ϕ is accessible via action a (or conversely, what would happen if the agent does a), but it does not answer the question of whether the agent will, in fact, do a.

We have used the notion of minimal change to reflect the context dependency of preference statements, but minimal change serves other purposes as well. Stalnaker introduced minimal change to modal logic in order to capture the semantics of the counterfactual conditionals that reflect causality [38,68]. Our update operator is a backward-looking dual to Stalnaker's conditional, as the "Ramsey rule" shows.

In our setup, the action operator is not bound to minimal change. Since actions entail causal effects, minimal change should appear in the semantics of the action operator [14,33,71]. This could address some nastier problems of action logics, in particular the *qualification, frame,* and *ramification* problem. Actions may require a specific context

[7] Contraposition has also been vigorously attacked with similar examples. The similarity is not surprising, since (CP) is, in fact, redundant; it follows from (CEP) and (N) as we have proved in Proposition 6.

for execution (qualification) that the action description must take into account. Linking actions to minimal change can provide an implicit qualification of the context of a specific action through the accessibility relation for that action. Using minimal change, this context is given by the actual state (that either will or will not permit action a to be executed); no additional specification of the context is required, once the accessibility relation for action a is given. The *frame of change* is given by those conditions that do not have to change as a function of a's execution. And the ramifications of an action are "automatically" captured by identifying the set of its weakest postconditions.

ALX does not put strong constraints on the closest world function. Stronger constraints might be desirable, perhaps even a full-fledged definition. We have refrained from defining the closest world function for ALX for two reasons. First, we wanted to provide a "logicians logic", i.e., a logic whose formal semantics is more than a faithful mirror of its syntax. As a consequence, we have used the standard semantic setup without strong restrictions on the definition of models and have not given a description of the properties of possible worlds. A definition of the closest world function would require such a description. Second, we are not sure about the exact meaning of the notion of "closest worlds", despite various attempts in the literature to provide a definition [14, 22, 23, 33, 36, 71]. There are two main problems: the definitions do not restrict the set of closest worlds as much as intuition seems to require (so the set contains more worlds than it should); also, the definitions do not clearly distinguish between epistemically closest worlds and causally closest worlds. This distinction is required, however, because the closest accessible world might very well be further away than the closest imaginable world, and this difference is important for an action logic.

We can illustrate this point by looking at minimal change as a consequence of an action. After all, the standard interpretation of actions is in terms of causality, hence in terms of minimal change. Unfortunately, there are several ways to conceptualize minimal change with respect to actions. We use $\langle a\rangle^{\#}\phi$ to denote the set of worlds where, by doing action a, the agent can achieve a minimally different ϕ-situation. One way to conceptualize such a change would be in terms of the closest world accessible via action a. Call this kind of "minimal change action" $\langle a\rangle^{\#1}$. The corresponding truth condition is:

$$[\![\langle a\rangle^{\#1}\phi]\!]_M = \{w : \exists w' \in W(w' \in cw(w, \{x : R^a wx\}) \text{ and } w' \in [\![\phi]\!]_M)\}.$$

According to this truth condition, $\langle a\rangle^{\#1}\phi$ first looks at the closest worlds accessible via action a and from this set picks the ϕ-worlds (see Fig. 3).

Consider, as an example, that a denotes the action of "slamming the door" and assume that *slamming* the door will cause the picture to fall off the wall (as opposed to, say, "closing the door" that will leave the picture unharmed). Assume ϕ stands for the fact that the door is shut. $\langle a\rangle^{\#1}\phi$ now looks at a world where the door is shut and the picture fell off the wall.

A second possible definition would approach the minimal change via minimally different ϕ-worlds. Call the corresponding minimal change action $\langle a\rangle^{\#2}$ (see Fig. 4). The corresponding truth condition is:

$$[\![\langle a\rangle^{\#2}\phi]\!]_M = \{w : \exists w' \in W(w' \in cw(w, [\![\phi]\!]_M) \text{ and } R^a ww')\}.$$

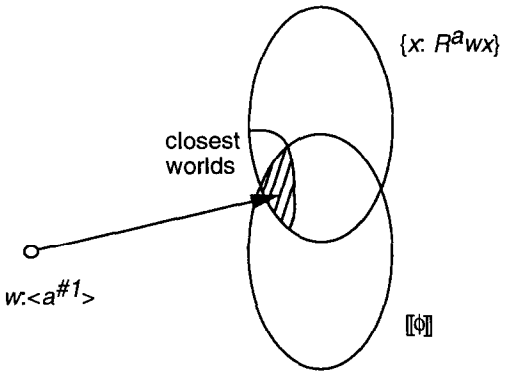

Fig. 3. Minimal change action $\langle a\rangle^{\#1}$.

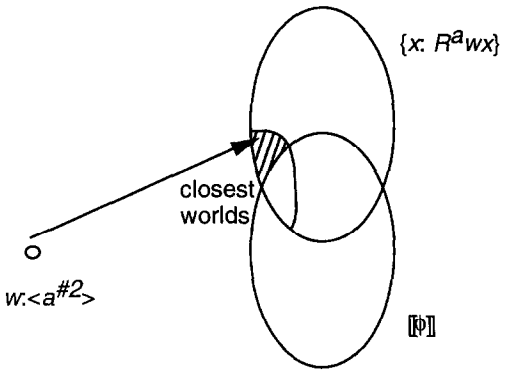

Fig. 4. Minimal change action $\langle a\rangle^{\#2}$.

Reconsider the previous example for $\langle a\rangle^{\#2}$. Slamming the door would now get us to worlds where the door is shut and the picture is back on the wall.

Since an additional action is implicit in $\langle a\rangle^{\#2}$, this kind of minimal change appears less intuitive than $\langle a\rangle^{\#1}$. But $\langle a\rangle^{\#1}$ has a drawback as well: $\langle a\rangle^{\#1}$ will give counterintuitive results if the intersection of accessible worlds and ϕ-worlds is not empty and the intersection of the closest a-accessible worlds with the ϕ-worlds is. This could happen, for example, if slamming the door would not shut the door (say, because of reverberation of the door frame). To cover this possibility we might want to look at the closest world in the intersection of a-accessible worlds and ϕ-worlds. On this view, the minimal change action is not going to return the empty set if the intersection of the accessible worlds and the closest worlds is not empty. Denote this kind of minimal change by $\langle a\rangle^{\#3}$ (see Fig. 5). The corresponding truth condition is:

$$[\![\langle a\rangle^{\#3}\phi]\!]_M = \{w : \exists w' \in W(w' \in cw(w, \{x : R^a wx\} \cap [\![\phi]\!]_M))\}.$$

But $\langle a\rangle^{\#3}$ cannot be the last word either, because it leaves undecided the question of whether the picture is on the wall or not. In sum, we should avoid a full-fledged definition of the closest world function until we can decide this—and possibly other—questions.

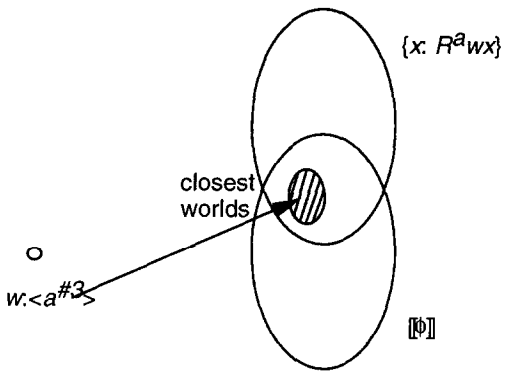

Fig. 5. Minimal change action $\langle a \rangle^{\#3}$.

6.3. Goals

Goal is an important primitive notion in other action logics [5,6,59], where the goal operators act as universal modalities. As a consequence, these logics have the necessitation rule for goals (if α is a theorem, then α is a goal) and the closure of goals under logical implication (if α is a goal and $\alpha \rightarrow \beta$ is a theorem, then β must be a goal). The necessitation rule and the deductive closure of goals have fairly severe counterintuitive implications. For example, if pain is always a consequence of having one's teeth fixed, then the pain itself becomes as a goal. Also, it does not make sense to treat tautologies as goals, as the necessitation rule would require. Much recent work in action logic has gone into systems that try to avoid these consequences by introducing an array of goal-related notions [5,6,59]. Unfortunately, these complications bring in other or additional counterintuitive effects of goals. For example, in Cohen and Levesque's logic [5,6], it is a theorem that if an agent believes that a fact holds, then the fact becomes the agent's goal. Rao and Georgeff's paper [59] avoids both necessitation and logical closure for certain epistemically qualified goals (agents need not adopt as goals what they *believe* to be *inevitably always* true and they need not to adopt ψ as a goal if they *believe* $\phi \rightarrow \psi$ to be *inevitably always* true and if they have ϕ as a goal). But in order to obtain these results, Rao and Georgeff have to make other counterintuitive assumptions. For example, they must assume that any believe-accessible world contains a goal. ALX can avoid both the necessitation rule and the deductive closure of goals by much simpler means, since we need not require monotonicity for the preference operator.

Proposition 22. (i) *Goals (as defined in this paper) are not closed under logical implication, i.e.,* $\models (\phi \rightarrow \psi)$ *does not imply* $\models (\boldsymbol{G}\phi \rightarrow \boldsymbol{G}\psi)$. (ii) *Furthermore, goals do not satisfy the necessitation rule, i.e.,* $\models \phi$ *does not imply* $\models \boldsymbol{G}\phi$.

Proof. (i) We construct a model $M = \langle W, cw, \succ, R^a, V \rangle$ for which $M \models (\phi \rightarrow \psi)$ and $M \not\models \boldsymbol{G}\phi \rightarrow \boldsymbol{G}\psi$ holds. Let $W = \{w1, w2\}$, let $cw : W \times \mathcal{P}(W) \rightarrow \mathcal{P}(W)$ be a function that satisfies (CS1)–(CSC). Let $\succ = \{\langle \{w1\}, \{w2\} \rangle\}$ and let R^a be any set. Define V

as follows: $V(p) = \{w1\}$, $V(q) = \{w1, w2\}$ and $V(r) = \{w2\}$.

It is easy to see that $[\![p]\!]_M \subset [\![q]\!]_M$, so that $M \models (p \rightarrow q)$. On the other hand, we have:

$$cw(w1, [\![p \wedge \neg r]\!]_M) = \{w1\}, \qquad cw(w1, [\![r \wedge \neg p]\!]_M) = \{w2\},$$
$$cw(w1, [\![q \wedge \neg r]\!]_M) = \{w1\}, \qquad cw(w1, [\![r \wedge \neg q]\!]_M) = \emptyset.$$

Therefore,

$$M, w1 \Vdash (p\boldsymbol{P}r) \wedge \neg(q\boldsymbol{P}r).$$

Hence,

$$M, w1 \Vdash \neg((p\boldsymbol{P}r) \rightarrow (q\boldsymbol{P}r)).$$

Thus the preference operator is not closed under logical implication. As a consequence, goals defined in terms of preferences are not closed under logical implication either.

(ii) We have shown that

$$\neg(\top \boldsymbol{P} \phi)$$

is a theorem of ALX, hence the preference operator does not satisfy the necessitation rule. □

6.4. Decision and planning

ALX is designed to represent theories about human actions, particularly theories about organizations. These theories are usually built around a decision cycle that has an agent pondering goals as a function of his problems and his action alternatives [35,54]. Planning, as understood in AI, is not a typical problem for such theories, because it is primarily a procedural problem: given a goal, find an an optimal sequence of actions and find it fast. ALX's logical properties (completeness, decidability) guarantee that an existence proof of an action sequence leading to a particular goal can always be found, provided this action sequence is feasible, i.e., a substructure of the accessibility relation R^a. This may be more than can be said of some other planners. As an example, consider the case of conjunctive planning discussed in [43].

We have a machine to buy cakes and apples; a cake costs a dollar and an apple three quarters. Due to an unfortunate design, the machine only accepts dollars and it returns a quarter when the user buys an apple; to alleviate in part this problem, the machine can change four quarters into a dollar (see Fig. 6).

One meaningful planning problem is: Assume a user has five quarters in his pocket: can he get a cake and have some change left? We can represent the domain as follows:

(1) having-one-dollar → [*buying-a-cake*]having-a-cake.

(2) having-one-dollar →
[*buying-an-apple*](having-an-apple ∧ having-one-quarter-left).

(3) having-four-quarters → [*change*]having-one-dollar.

(4) having-five-quarters → having-four-quarters ∧ having-one-quarter-left.

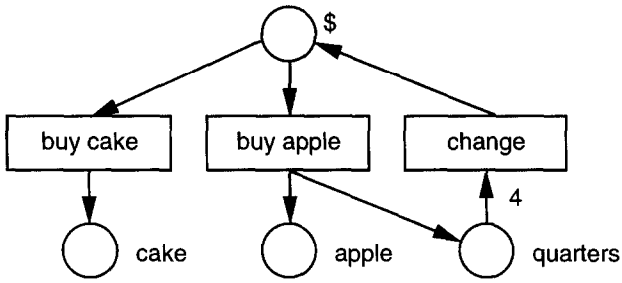

Fig. 6. A simple conjunctive planning problem.

In the next three assumptions, we exploit the fact that the universal action modality can be void:

(5) having-one-quarter-left → [*buying-a-cake*]having-one-quarter-left.

(6) having-one-quarter-left → [*buying-an-apple*]having-one-quarter-left.

(7) having-one-quarter-left → [*change*]having-one-quarter-left.

(8) having-one-quarter-left → having-some-quarter-left.

The following statement describes the specific situation of the user:

(9) having-five-quarters.

The planning problem is whether or not there exists an action sequence that makes the following state accessible:

(having-a-cake ∧ having-some-change-left).

In proving this state from the premises, we generate the required action sequence: the proof is as follows:[8]

	having-five-quarters	(9)
⇒	having-four-quarters ∧ having-one-quarter-left	(4)
⇒	having-one-quarter-left ∧ [*change*]having-one-dollar	(3)
⇒	having-one-quarter-left ∧ [*change*][*buying-a-cake*]having-a-cake	(1, MONA)

[8] In the proof, we use the ALX theorem $[a](\phi \wedge \psi) \leftrightarrow [a]\phi \wedge [a]\psi$, which is derivable from ALXS, since:

	$\langle a\rangle(\neg\phi \vee \neg\psi) \leftrightarrow \langle a\rangle\neg\phi \vee \langle a\rangle\neg\psi$	(A2)
⇔	$\neg\langle a\rangle(\neg\phi \vee \neg\psi) \leftrightarrow \neg(\langle a\rangle\neg\phi \vee \langle a\rangle\neg\psi)$	(Meta-reasoning)
⇔	$\neg\langle a\rangle\neg(\phi \wedge \psi) \leftrightarrow \neg\langle a\rangle\neg\phi \wedge \neg\langle a\rangle\neg\psi$	(Meta-reasoning)
⇔	$[a](\phi \wedge \psi) \leftrightarrow [a]\phi \wedge [a]\psi$	(Definition of [ ]).

$\Rightarrow$ [*change*]having-one-quarter-left $\wedge$
[*change*][*buying-a-cake*]having-a-cake (7)
$\Rightarrow$ [*change*][*buying-a-cake*]having-one-quarter-left $\wedge$
[*change*][*buying-a-cake*]having-a-cake (5, MONA)
$\Rightarrow$ [*change*][*buying-a-cake*]having-some-quarter-left $\wedge$
[change][*buying-a-cake*]having-a-cake (8, MONA)
$\Rightarrow$ [*change*][*buying-a-cake*]
(having-some-quarter-left $\wedge$ having-a-cake) (ALX theorem)
$\Rightarrow$ [*change*][*buying-a-cake*]
(having-a-cake $\wedge$ having-some-quarter-left) (SUBA).

ALX has no machinery for finding plans efficiently. Also, ALX has, in its present version, no way of linking actions (or, for that matter, attitudes) to the execution of plans; its action modalities are contemplative, so there is no direct way to express that a decision or an action has, in fact, occurred. An indirect way to simulate decisions would be by defining a necessity operator[9] and saying that choosing a particular goal makes this goal necessary

$$G\phi \rightarrow \Box G\phi.$$

But this expression does not fully capture the intuitive meaning of "decision". A *do/has-done* operator would be helpful, but such an operator requires a semantic setup that includes time explicitly.

ALX could represent a planning procedure that builds an evaluation gradually as the planning process progresses, simply by conditioning preferences on states, so that certain states allow for an evaluation whereas others don't. But, again, this representation would not cover the dynamic flavor of a real planning process where "the world out there" acts as an oracle and the task is to make this oracle talk and talk fast.

6.5. Expressive power

There are still many limitations to ALX's expressive power. The most important, we think, is ALX's present restriction to a propositional description language; ALX's construction suggests a straightforward extension to first-order logic. A second important limitation is due to the absence of a belief operator. We cannot distinguish between "objectively" available knowledge and the knowledge available to an individual decision maker, so we cannot model the difference between objective and individual knowledge. For example, we cannot distinguish between disbelief in inaccessibility and accessibility, although this distinction is important when a rational agent is pondering its goals. Third, we may need time operators to represent the process of causality or to define a do-operator. For example, a time operator would be useful to express a notion of "tradition", or of expectations regarding the future. Last, but not least, we may want to extend ALX

[9] This can be done as follows: $\Box\phi \stackrel{\text{def}}{\Longleftrightarrow} \neg\phi \rightsquigarrow \phi$.

to multi-agent acting, by allowing for indexing modal operators with agent terms and for quantification over agents and actions.

We have tried to incorporate important elements of bounded rationality into ALX. The basic message of bounded rationality is quite simple: remember the limits of human information-processing capacity. Yet it is one thing to recognize the abstract existence of these limits and another one to find out where these limits are drawn. In the first case, one has to make sure that omniscience claims or omnipotence claims are avoided. In the second case, one has to identify which information is processed and how. No logic would be able to fully answer the second question, since it is to a large extent an empirical one. However, the question does have some general aspects that we did not address in this paper. It was not very difficult to transpose the declarative aspects of Simon's original conceptualization of bounded rationality into an action logic. However, its procedural aspects, especially search, may require a completely different semantic semantic setup. Incorporating search explicitly in the logic (rather than just naming it, as we did in the definition of satisficing states) seems to require introducing "information" as a distinct object to the logic. This, in turn, seems to require at least a partial logic; perhaps that future work may be able to exploit the progress of situation semantics in this area [57,58].

7. Conclusions and future direction

ALX is the first preference-based action logic. Its basic construction is fairly simple: let an agent have context-dependent preferences, give it action alternatives and let it deliberate about its actions on the basis of preferences and action alternatives. Furthermore, add a notion of causality that reflects the idea of minimal change. The context-dependent construction of preferences and the ability to build nested preferences gives ALX a strong expressive potential—although realizing this potential will require an extension of the logic to a multi-agent setup and to a first-order description language. Also, ALX's basic construction avoids important weaknesses of other action and preference logics. For example, goals in ALX need not be closed under logical implication, nor are tautologies automatically goals. Furthermore, the infamous conjunction expansion principle (CEP) from von Wright's logic is tamed by normality (N). ALX is complete and decidable, which prepares the ground for the development of an ALX-theorem prover. Furthermore, its pleasant logical properties would even allow for the use of ALX as a planner, albeit an inefficient one. However, ALX's primary task is knowledge representation. Preliminary experiments with a first-order version of ALX show that the combination of preference, action and update modalities allows for a flexible representation of a variety of theoretical problems in organization theory [46]. Our plan is to build a sequence of more expressive versions of ALX, starting with an extension to first-order logic as a description language, then adding multi-agent facilities, a belief operator, time operators and (if possible) a "do" operator. Work on ALX will continue.

Appendix A. Proofs

A.1. More properties of the preference operator

Proposition 6 (More properties of the preference operator). *The following formulas are theorems of ALX*:

(CP) $\phi \boldsymbol{P} \psi \leftrightarrow (\neg\psi)\boldsymbol{P}(\neg\phi)$,
(IR) $\neg(\phi \boldsymbol{P} \phi)$,
(NT) $\neg(\top \boldsymbol{P} \phi), \neg(\phi \boldsymbol{P} \top)$,
(AS) $\phi \boldsymbol{P} \psi \rightarrow \neg(\psi \boldsymbol{P} \phi)$.

Proof.

(CP) $\phi \boldsymbol{P} \psi \leftrightarrow (\neg\psi)\boldsymbol{P}(\neg\phi)$.

$\vdash \phi \boldsymbol{P} \psi$
$\Leftrightarrow \vdash (\phi \wedge \neg\psi)\boldsymbol{P}(\psi \wedge \neg\phi)$ (CEP)
$\Leftrightarrow \vdash (\neg\psi \wedge \neg(\neg\phi))\boldsymbol{P}(\neg\phi \wedge (\neg(\neg\psi)))$ (SUBP)
$\Leftrightarrow \vdash \neg\psi \boldsymbol{P} \neg\phi$ (CEP)

(IR) $\neg(\phi \boldsymbol{P} \phi)$.

$\vdash (\phi \boldsymbol{P} \phi)$
$\Rightarrow \vdash (\phi \wedge \neg\phi)\boldsymbol{P}(\phi \wedge \neg\phi)$ (CEP)
$\Rightarrow \vdash \bot \boldsymbol{P} \bot$ (Definition of $\bot$)
$\Rightarrow \vdash \bot$ (N)

So $\vdash (\phi \boldsymbol{P} \phi) \rightarrow \bot$. Therefore, $\vdash \neg(\phi \boldsymbol{P} \phi)$.

(NT) $\neg(\top \boldsymbol{P} \phi)$, $\neg(\phi \boldsymbol{P} \top)$.

$\vdash (\top \boldsymbol{P} \phi)$
$\Rightarrow \vdash (\neg\phi)\boldsymbol{P}(\neg\top)$ (CP)
$\Rightarrow \vdash (\neg\phi)\boldsymbol{P}\bot$ (Definition of $\top$)
$\Rightarrow \vdash \bot$ (N)

So, $\vdash \neg(\top \boldsymbol{P} \phi)$. The proof for the second half of (NT) is similar.

(AS) $\phi \boldsymbol{P} \psi \rightarrow \neg(\psi \boldsymbol{P} \phi)$.

$\vdash (\phi \boldsymbol{P} \psi) \wedge (\psi \boldsymbol{P} \phi)$
$\Rightarrow \vdash \phi \boldsymbol{P} \phi$ (TR)
$\Rightarrow \vdash \bot$ (IR) □

A.2. Soundness

Lemma 23 (Soundness of the update axioms). (U1)–(U6) *are valid on the class of ALX models.*

Proof.

(U1) $\phi \circ \psi \rightarrow \psi$.

$$
\begin{aligned}
& M, w \Vdash \phi \circ \psi \\
\Leftrightarrow\ & \exists i \in [\![\phi]\!]_M (w \in cw(i, [\![\psi]\!]_M)) && \text{(Truth condition)} \\
\Rightarrow\ & \exists i \in [\![\phi]\!]_M (w \in [\![\psi]\!]_M) && \text{(CS1)} \\
\Rightarrow\ & w \in [\![\psi]\!]_M && \text{(Meta-reasoning)} \\
\Leftrightarrow\ & M, w \Vdash \psi && \text{(Definition of } \Vdash\text{)}
\end{aligned}
$$

(U2) $\phi \wedge \psi \rightarrow \phi \circ \psi$.

$$
\begin{aligned}
& M, w \Vdash \phi \wedge \psi \\
\Leftrightarrow\ & M, w \Vdash \phi \text{ and } M, w \Vdash \psi && \text{(Truth condition)} \\
\Leftrightarrow\ & w \in [\![\phi]\!]_M \text{ and } w \in [\![\psi]\!]_M && \text{(Definition of } \Vdash\text{)} \\
\Rightarrow\ & w \subset [\![\phi]\!]_M \text{ and } cw(w, [\![\psi]\!]_M) = \{w\} && \text{(CS2)} \\
\Rightarrow\ & \exists w (w \in [\![\phi]\!]_M \text{ and } w \in cw(w, [\![\psi]\!]_M)) && \text{(Meta-reasoning)} \\
\Leftrightarrow\ & M, w \Vdash \phi \circ \psi && \text{(Truth condition)}
\end{aligned}
$$

(U3) $\neg(\bot \circ \phi), \neg(\phi \circ \bot)$.

$$
\begin{aligned}
& M, w \Vdash \bot \circ \phi \\
\Leftrightarrow\ & \exists i (i \in [\![\bot]\!]_M \text{ and } w \in cw(i, [\![\phi]\!]_M)) && \text{(Truth condition)} \\
\Rightarrow\ & \exists i (i \in [\![\bot]\!]_M) && \text{(Meta-reasoning)} \\
\Rightarrow\ & \mathit{False} && \text{(Meta-reasoning)}
\end{aligned}
$$

(U4) $(\phi \vee \psi) \circ \chi \leftrightarrow (\phi \circ \chi) \vee (\psi \circ \chi)$.

$$
\begin{aligned}
& M, w \Vdash (\phi \vee \psi) \circ \chi \\
\Leftrightarrow\ & \exists i (i \in [\![\phi \vee \psi]\!]_M \text{ and } w \in cw(i, [\![\chi]\!]_M)) && \text{(Truth condition)} \\
\Leftrightarrow\ & \exists i ((i \in [\![\phi]\!]_M \text{ and } w \in cw(i, [\![\chi]\!]_M)) \text{ or} \\
& \qquad (i \in [\![\psi]\!] \text{ and } w \in cw(i, [\![\chi]\!]_M))) && \text{(Meta-reasoning)} \\
\Leftrightarrow\ & w \Vdash (\phi \circ \chi) \text{ or } w \Vdash (\psi \circ \chi) && \text{(Truth condition)} \\
\Leftrightarrow\ & w \Vdash (\phi \circ \chi) \vee (\psi \circ \chi) && \text{(Truth condition)}
\end{aligned}
$$

(U5) $(\phi \wedge \psi) \circ \psi \rightarrow \phi$.

$$
\begin{aligned}
& M, w \Vdash (\phi \wedge \psi) \circ \psi \\
\Leftrightarrow\ & \exists i (i \in [\![\phi \wedge \psi]\!]_M \text{ and } w \in cw(i, [\![\psi]\!]_M)) && \text{(Truth condition)} \\
\Rightarrow\ & \exists i (i \in [\![\phi]\!]_M \text{ and } i \in [\![\psi]\!]_M \text{ and } w \in cw(i, [\![\psi]\!]_M)) && \text{(Truth condition)} \\
\Rightarrow\ & \exists i (i \in [\![\phi]\!]_M \text{ and } w = i) && \text{(CS2)} \\
\Rightarrow\ & w \in [\![\phi]\!]_M && \text{(Meta-reasoning)} \\
\Rightarrow\ & M, w \Vdash [\![\phi]\!]_M && \text{(Definition of } \Vdash\text{)}
\end{aligned}
$$

(U6) $(\phi \circ \psi) \wedge \chi \rightarrow \phi \circ (\psi \wedge \chi)$.

$$
\begin{aligned}
& M, w \Vdash (\phi \circ \psi) \wedge \chi \\
\Leftrightarrow\ & \exists i (i \in [\![\phi]\!]_M \text{ and } w \in cw(i, [\![\psi]\!]_M)) \text{ and } w \in [\![\chi]\!]_M && \text{(Truth condition)} \\
\Rightarrow\ & \exists i (i \in [\![\phi]\!]_M \text{ and } w \in cw(i, [\![\psi]\!]_M) \cap [\![\chi]\!]_M)) && \text{(Meta-reasoning)} \\
\Rightarrow\ & \exists i (i \in [\![\phi]\!]_M \text{ and } w \in cw(i, [\![\psi \wedge \chi]\!]_M) && \text{(CS4)} \\
\Leftrightarrow\ & M, w \Vdash \phi \circ (\psi \wedge \chi) && \text{(Truth condition)} \quad \square
\end{aligned}
$$

Lemma 24 (Soundness of the preference axioms). (CEP), (TR), *and* (N) *are valid on the class of ALX models.*

Proof. For any ALX model $M = \langle W, cw, \succ, R_a, V \rangle$ and any $w \in W$:

(CEP) $\phi \boldsymbol{P} \psi \leftrightarrow (\phi \wedge \neg\psi) \boldsymbol{P} (\psi \wedge \neg\phi)$.

$$
\begin{aligned}
& M, w \Vdash \phi \boldsymbol{P} \psi \\
\Leftrightarrow\ & cw(w, [\![\phi \wedge \neg\psi]\!]_M) \succ cw(w, [\![\psi \wedge \neg\phi]\!]_M) && \text{(Truth condition)} \\
\Leftrightarrow\ & cw(w, [\![(\phi \wedge \neg\psi) \wedge \neg(\psi \wedge \neg\phi)]\!]_M) \succ \\
& \quad cw(w, [\![(\psi \wedge \neg\phi) \wedge \neg(\phi \wedge \neg\psi)]\!]_M) && \text{(Propositional logic)} \\
\Leftrightarrow\ & M, w \Vdash (\phi \wedge \neg\psi) \boldsymbol{P} (\psi \wedge \neg\phi) && \text{(Truth condition)}
\end{aligned}
$$

(TR) $(\phi \boldsymbol{P} \psi) \wedge (\psi \boldsymbol{P} \chi) \rightarrow (\phi \boldsymbol{P} \chi)$.

$$
\begin{aligned}
& M, w \Vdash (\phi \boldsymbol{P} \psi) \wedge (\psi \boldsymbol{P} \chi) \\
\Leftrightarrow\ & cw(w, [\![\phi \wedge \neg\psi]\!]_M) \succ cw(w, [\![\psi \wedge \neg\phi]\!]_M) \text{ and} \\
& \quad cw(w, [\![\psi \wedge \neg\chi]\!]_M) \succ cw(w, [\![\chi \wedge \neg\psi]\!]_M) && \text{(Truth condition)} \\
\Rightarrow\ & cw(w, [\![\phi \wedge \neg\chi]\!]_M) \succ cw(w, [\![\chi \wedge \neg\phi)]\!]_M) && \text{(TRAN)} \\
\Leftrightarrow\ & M, w \Vdash (\phi \boldsymbol{P} \chi) && \text{(Truth condition)}
\end{aligned}
$$

(N) $\neg(\bot \boldsymbol{P} \phi), \neg(\phi \boldsymbol{P} \bot)$.

$$
\begin{aligned}
& M, w \Vdash \bot \boldsymbol{P} \phi \\
\Leftrightarrow\ & cw(w, [\![\bot \wedge \neg\phi]\!]_M) \succ cw(w, [\![\phi \wedge \top]\!]_M) && \text{(Truth condition)} \\
\Rightarrow\ & cw(w, \emptyset) \succ cw(w, [\![\phi]\!]_M) && \text{(Propositional logic)} \\
\Rightarrow\ & \emptyset \succ cw(w, [\![\phi]\!]_M) && \text{(CS1)} \\
\Rightarrow\ & \mathit{False} && \text{(NORM of } \succ)
\end{aligned}
$$

The proof about $\neg(\phi \boldsymbol{P} \bot)$ goes symmetrically. $\square$

Lemma 25 (Soundness of the action axioms). (A1)–(A3) *are valid on the class of ALX models.*

Proof. For any ALX model $M = \langle W, cw, \succ, R_a, V \rangle$ and any $w \in W$:

(A1) $\langle a\rangle\bot \leftrightarrow \bot$.

$$
\begin{aligned}
& M, w \Vdash \langle a\rangle\bot \\
\Leftrightarrow\ & \exists z(R^a wz \text{ and } z \in [\![\bot]\!]_M) && \text{(Truth condition)} \\
\Rightarrow\ & \exists z(z \in \emptyset) && \text{(Meta-reasoning)} \\
\Rightarrow\ & \mathit{False} && \text{(Meta-reasoning)}
\end{aligned}
$$

By propositional logic, $\bot \rightarrow \langle a\rangle\bot$. So $\langle a\rangle\bot \leftrightarrow \bot$.

(A2) $\langle a\rangle(\phi \vee \psi) \leftrightarrow \langle a\rangle\phi \vee \langle a\rangle\psi$.

$$
\begin{aligned}
& M, w \Vdash \langle a\rangle(\phi \vee \psi) \\
\Leftrightarrow\ & \exists z(R^a wz \text{ and } (z \in [\![\phi]\!]_M \text{ or } z \in [\![\psi]\!]_M)) && \text{(Truth condition)} \\
\Leftrightarrow\ & \exists z(R^a wz \text{ and } z \in [\![\phi]\!]_M) \text{ or } \exists z(R^a wz \text{ and } z \in [\![\psi]\!]_M) && \text{(Meta-reasoning)} \\
\Leftrightarrow\ & M, w \Vdash (\langle a\rangle\phi \vee \langle a\rangle\psi) && \text{(Truth condition)}
\end{aligned}
$$

(A3) $\langle a\rangle(\phi \wedge \psi) \rightarrow (\langle a\rangle\phi \wedge \langle a\rangle\psi)$.

$$
\begin{aligned}
& M, w \Vdash \langle a\rangle(\phi \wedge \psi) \\
\Leftrightarrow\ & \exists z(R^a wz \text{ and } z \in [\![\phi \wedge \psi]\!]_M) && \text{(Truth condition)} \\
\Leftrightarrow\ & \exists z(R^a wz \text{ and } z \in [\![\phi]\!]_M \text{ and } z \in [\![\psi]\!]_M) && \text{(Truth condition)} \\
\Rightarrow\ & \exists z(R^a wz \text{ and } z \in [\![\phi]\!]_M) \text{ and} \\
& \exists z(R^a wz \text{ and } z \in [\![\psi]\!]_M) && \text{(Meta-reasoning)} \\
\Leftrightarrow\ & M, w \Vdash \langle a\rangle\phi \text{ and } M, w \Vdash \langle a\rangle\psi && \text{(Truth condition)} \\
\Leftrightarrow\ & M, w \Vdash (\langle a\rangle\phi \wedge \langle a\rangle\psi) && \text{(Truth condition)} \quad \square
\end{aligned}
$$

Lemma 26 (Soundness of the inference rules). (MP), (SUBA), (SUBU), (SUBP), (NECA), (MONA) *and* (MONU) *are validity-preserving for the class of ALX models.*

Proof. For any ALX model $M = \langle W, cw, \succ, R^a, V\rangle$:

(MP) $\vdash \phi$, $\vdash (\phi \rightarrow \psi) \Rightarrow \vdash \psi$.

$$
\begin{aligned}
& \phi \text{ and } (\phi \rightarrow \psi) \text{ are valid for } M \\
\Rightarrow\ & \forall w \in W(M, w \Vdash \phi) \text{ and} \\
& \forall w \in W(M, w \Vdash (\phi \rightarrow \psi)) && \text{(Definition of validity)} \\
\Rightarrow\ & \forall w \in W(M, w \Vdash \phi \text{ and } M, w \Vdash (\phi \rightarrow \psi)) && \text{(Meta-reasoning)} \\
\Rightarrow\ & \forall w \in W(M, w \Vdash \phi \wedge (\phi \rightarrow \psi)) && \text{(Truth condition)} \\
\Rightarrow\ & \forall w \in W(M, w \Vdash \psi) && \text{(Definition of } \rightarrow\text{)} \\
\Rightarrow\ & \psi \text{ is valid for } M && \text{(Definition of validity)}
\end{aligned}
$$

(SUBA) $\vdash (\phi \leftrightarrow \phi') \Rightarrow \vdash (\langle a\rangle\phi \leftrightarrow \langle a\rangle\phi')$.

$\phi \leftrightarrow \phi'$ and $\langle a\rangle\phi$ are valid for M
$\Rightarrow \forall w \in W(M, w \Vdash (\phi \leftrightarrow \phi'))$ and
$\forall w \in W(M, w \Vdash \langle a\rangle\phi)$ (Definition of validity)
$\Rightarrow \forall w \in W(M, w \Vdash \phi \leftrightarrow \phi')$ and
$\forall w \in W(\exists w' \in W)(R^a w w'$ and $M, w' \Vdash \phi))$ (Truth condition)
$\Rightarrow (\forall w \in W)(\exists w' \in W)(R^a w w'$ and $M, w' \Vdash \phi')$ (Meta-reasoning)
$\Rightarrow (\forall w \in W)(M, w \Vdash \langle a\rangle\phi')$ (Truth condition)
$\Rightarrow \langle a\rangle\phi$ is valid for M (Definition of validity)

Therefore, $\vdash \phi \leftrightarrow \phi' \Rightarrow \vdash \langle a\rangle\phi \rightarrow \langle a\rangle\phi'$ is validity-preserving on a model. Symmetrically, we can show that $\vdash \phi \leftrightarrow \phi' \Rightarrow \vdash \langle a\rangle\phi' \rightarrow \langle a\rangle\phi$ is validity-preserving. So (SUBA) is sound.

The soundness of (SUMP) and (SUMU) is established in a similar fashion.

(NECA) $\vdash \phi \Rightarrow \vdash [a]\phi$.

ϕ is valid for M
$\Rightarrow (\forall w \in W)(M, w \Vdash \phi)$ (Definition of validity)
$\Rightarrow (\forall w \in W)(\forall w' \in W)(R^a w w' \Rightarrow M, w' \Vdash \phi)$ (Meta-reasoning)
$\Rightarrow (\forall w \in W)(M, w \Vdash [a]\phi)$ (Truth condition)
$\Rightarrow [a]\phi$ is valid for M. (Definition of validity)

(MONA) $\vdash \langle a\rangle\phi, \vdash (\phi \rightarrow \psi) \Rightarrow \vdash \langle a\rangle\psi$.

$\langle a\rangle\phi$ and $\phi \rightarrow \psi$ are valid for M
$\Rightarrow \forall w \in W(M, w \Vdash \langle a\rangle\phi)$ and $[\![\phi]\!]_M \subseteq [\![\psi]\!]_M$ (Definition of validity)
$\Rightarrow \forall w \in W(\exists w'(R^a w w'$ and $w' \in [\![\phi]\!])$ and
$[\![\phi]\!]_M \subseteq [\![\psi]\!]_M$ (Truth condition)
$\Rightarrow \forall w \in W(\exists w'(R^a w w'$ and $w' \in [\![\psi]\!]_M)$ (Meta-reasoning)
$\Rightarrow \forall w \in W(M, w \Vdash \langle a\rangle\psi)$ (Truth condition)
$\Rightarrow \langle a\rangle\psi$ is valid for M (Definition of validity)

(MONU) $\vdash \phi \circ \chi, \vdash (\phi \rightarrow \psi) \Rightarrow \vdash \psi \circ \chi$.

$\phi \circ \chi$ and $\phi \rightarrow \psi$ are valid for M
$\Rightarrow \forall w \in W(M, w \Vdash \phi \circ \chi)$ and $[\![\phi]\!]_M \subseteq [\![\psi]\!]_M$ (Definition of validity)
$\Rightarrow \forall w \in W(\exists w'(w' \in [\![\phi]\!]$ and
$w \in cw(w', [\![\chi]\!]_M)$ and
$[\![\phi]\!] \subseteq [\![\psi]\!]_M)$ (Truth condition)
$\Rightarrow \forall w \in W(\exists w'(w' \in [\![\psi]\!]_M$ and $w \in cw(w', [\![\chi]\!]_M))$ (Meta-reasoning)
$\Rightarrow \forall w \in W(M, w \Vdash \psi \circ \chi)$ (Truth condition)
$\Rightarrow \psi \circ \chi$ is valid for M (Definition of validity) □

Proposition 7 (Soundness of ALXS). *ALXS is sound.*

Proof. Lemmas A.1–A.4 together imply it. □

A.3. More properties of the update operator

Proposition 27 (Update theorems). *The following propositions are sound for the class of ALX models:*

(U1°) $\phi \circ \phi \rightarrow \phi$.
(U2°) $(\phi \circ \psi) \circ \psi) \rightarrow \phi \circ \psi$.
(U3°) $\neg\phi \wedge (\phi \circ \psi) \rightarrow (\phi \wedge \neg\psi) \circ \psi$.
(U4°) $(\phi \circ \psi \rightarrow \bot) \rightarrow ((\phi \rightarrow \bot) \vee (\psi \rightarrow \bot))$.
(U5°) $(\phi \circ \psi) \wedge \neg\phi \rightarrow \neg\psi \circ \psi$.
(U6°) $(\neg\phi \circ \psi) \wedge (\neg\psi \circ \phi) \rightarrow (\neg\phi \circ \phi) \wedge (\neg\psi \circ \psi)$.
(U7°) $(\phi \wedge \psi) \circ \phi \leftrightarrow (\phi \wedge \psi) \circ \psi$.
(U8°) $((\rho \wedge \phi) \circ \phi) \wedge \psi \rightarrow (\rho \wedge \psi) \circ \phi$.

Proof.

(U1°) $\phi \circ \phi \rightarrow \phi$.

$M, w \Vdash (\phi \circ \phi)$
$\Leftrightarrow \exists i(i \in [\![\phi]\!]_M$ and $w \in cw(i, [\![\phi]\!]_M))$ (Truth condition)
$\Leftrightarrow w \in [\![\phi]\!]_M$ (CS2)
$\Leftrightarrow M, w \Vdash \phi$ (Definition of $\Vdash$)

(U2°) $(\phi \circ \psi) \circ \psi \leftrightarrow \phi \circ \psi$. First we prove ($\Rightarrow$).

$M, w \Vdash (\phi \circ \psi) \circ \psi$
$\Leftrightarrow \exists i(i \in [\![\phi \circ \psi]\!]_M$ and $w \in cw(i, [\![\psi]\!]_M))$ (Truth condition)
$\Leftrightarrow \exists i \exists j(j \in [\![\phi]\!]_M$ and $i \in cw(j, [\![\psi]\!]_M)$ and
$w \in cw(i, [\![\psi]\!]_M))$ (Truth condition)
$\Rightarrow \exists i \exists j(j \in [\![\phi]\!]_M$ and $i \in cw(j, [\![\psi]\!]_M)$ and $w = i)$ (CS2)
$\Leftrightarrow \exists j(j \in [\![\phi]\!]_M$ and $w \in cw(j, [\![\psi]\!]_M))$ (Meta-reasoning)
$\Leftrightarrow M, w \Vdash \phi \circ \psi$ (Truth condition)

Next we prove ($\Leftarrow$).

$M, w \Vdash (\phi \circ \psi)$
$\Rightarrow M, w \Vdash (\phi \circ \psi) \wedge \psi$ (U1)
$\Rightarrow M, w \Vdash (\phi \circ \psi) \circ \psi$ (U2)

Therefore, $(\phi \circ \psi) \leftrightarrow (\phi \circ \psi) \circ \psi$.

(U3°) $\neg\phi \wedge (\phi \circ \psi) \rightarrow (\phi \wedge \neg\psi) \circ \psi$.

$M, w \Vdash \neg\phi \wedge (\phi \circ \psi)$
$\Leftrightarrow M, w \Vdash \neg\phi$ and $M, w \Vdash \phi \circ \psi$ (Truth condition)
$\Leftrightarrow M, w \Vdash \neg\phi$ and $\exists i((i \in [\![\phi]\!]_M)$ and $w \in cw(i, [\![\psi]\!]_M))$ (Truth condition)

Case 1: $i \in [\![\neg\psi]\!]_M$.

$i \in [\![\neg\psi]\!]_M$
$\Rightarrow \exists i(i \in [\![\phi]\!]_M$ and $i \in [\![\neg\psi]\!]_M$ and $w \in cw(i, [\![\psi]\!]_M))$ (Assumption)
$\Leftrightarrow \exists i(i \in [\![\phi \wedge \neg\psi]\!]_M$ and $w \in cw(i, [\![\psi]\!]_M))$ (Truth condition)
$\Leftrightarrow M, w \Vdash (\phi \wedge \neg\psi) \circ \psi$ (Truth condition)

Case 2: $i \in [\![\psi]\!]_M$.

$i \in [\![\psi]\!]_M$
$\Rightarrow i = w$ ($w \in cw(i, [\![\psi]\!]_M)$ and (CS2))
$\Rightarrow M, w \Vdash \phi$ and $M, w \Vdash \neg\phi$ ($i \in [\![\phi]\!]_M$ and $M, w \Vdash \neg\phi$)
$\Rightarrow$ *False*

(U4°) $(\phi \circ \psi \rightarrow \bot) \rightarrow ((\phi \rightarrow \bot) \vee (\psi \rightarrow \bot))$.

$(\phi \wedge \psi) \rightarrow (\phi \circ \psi)$ is valid
$\Leftrightarrow \neg(\phi \circ \psi) \rightarrow \neg\phi \vee \neg\psi$ is valid
$\Leftrightarrow (\neg(\phi \circ \psi) \vee \bot) \rightarrow (\neg\phi \vee \bot) \vee (\neg\psi \vee \bot)$ is valid
$\Leftrightarrow (\phi \circ \psi \rightarrow \bot) \rightarrow (\phi \rightarrow \bot) \vee (\psi \rightarrow \bot)$ is valid

(U5°) $(\phi \circ \psi) \wedge \neg\phi \rightarrow \neg\psi \circ \psi$.

$M, w \Vdash (\phi \circ \psi) \wedge \neg\phi$
$\Leftrightarrow \exists i(i \in [\![\phi]\!]_M$ and $w \in cw(i, [\![\psi]\!]_M)$ and $w \in [\![\neg\phi]\!]_M)$ (Truth condition)
$\Rightarrow \exists i \neq w$ (Meta-reasoning)

Suppose that $M, i \Vdash \psi$, then

$M, i \Vdash \psi \Rightarrow cw(i, [\![\psi]\!]_M) = \{i\} \Rightarrow w = i \Rightarrow$ *False*.

Therefore, $M, i \nVdash \psi$, namely, $M, w \Vdash \neg\psi \circ \psi$.

(U6°) $(\neg\phi \circ \psi) \wedge (\neg\psi \circ \phi) \rightarrow (\neg\phi \circ \phi) \wedge (\neg\psi \circ \psi)$.

$M, w \Vdash (\neg\psi \circ \phi) \wedge (\neg\phi \circ \psi)$
$\Rightarrow M, w \Vdash (\neg\psi \circ \phi) \wedge \psi \wedge (\neg\phi \circ \psi) \wedge \phi)$ (U1)
$\Rightarrow M, w \Vdash (\neg\psi \circ \psi) \wedge (\neg\phi \circ \phi)$ (U5°)

(U7°) $(\phi \wedge \psi) \circ \phi \leftrightarrow (\phi \wedge \psi) \circ \psi$.

$$
\begin{aligned}
& M, w \Vdash (\phi \wedge \psi) \circ \phi \\
\Rightarrow\ & M, w \Vdash ((\phi \wedge \psi) \circ \phi) \wedge \phi && \text{(U1)} \\
\Rightarrow\ & M, w \Vdash (\psi \wedge \phi) && \text{(U5)} \\
\Rightarrow\ & M, w \Vdash (\psi \wedge \phi) \circ \psi && \text{(U2)}
\end{aligned}
$$

Therefore, $(\phi \wedge \psi) \circ \phi \rightarrow (\phi \wedge \psi) \circ \psi$. The proof about $(\phi \wedge \psi) \circ \psi \rightarrow (\phi \wedge \psi) \circ \phi$ goes analogously.

(U8°) $((\rho \wedge \phi) \circ \phi) \wedge \psi \rightarrow (\rho \wedge \psi) \circ \phi$.

$$
\begin{aligned}
& M, w \Vdash ((\rho \wedge \phi) \circ \phi) \wedge \psi \\
\Rightarrow\ & M, w \Vdash \rho \wedge ((\rho \wedge \phi) \circ \phi) \wedge \psi && \text{(U5)} \\
\Rightarrow\ & M, w \Vdash \rho \wedge \phi \wedge \psi && \text{(U1)} \\
\Rightarrow\ & M, w \Vdash (\rho \wedge \psi) \circ \phi && \text{(U2)} \quad \square
\end{aligned}
$$

A.4. Completeness

The completeness proof for ALX proceeds along the lines of a Henkin-style construction. We give a detailed proof. First, we need a definition of consistency. We say that a formula φ is *consistent* (with respect to an axiom system) if $\neg\varphi$ is not provable (from that axiom system). A finite set $\{\varphi_1, \ldots, \varphi_k\}$ is consistent exactly if the formula $\varphi_1 \wedge \cdots \wedge \varphi_k$ is consistent. An infinite set of formulas is consistent if every finite subset of it is consistent. A set F of formulas is a *maximal consistent set* if it is consistent and any strict superset is inconsistent. With standard techniques of propositional reasoning it can be shown:

Lemma 28 (Lindenbaum's Lemma). *In any axiom system that includes all tautologies of propositional logic and the inference rule* (MP):

(1) *Any consistent set can be extended to a maximal consistent set.*
(2) *If F is a maximal consistent set, then for all formulas φ and ψ:*
  (a) *either $\varphi \in F$ or $\neg\varphi \in F$,*
  (b) *$\varphi \wedge \psi \in F$ iff $\varphi \in F$ and $\psi \in F$,*
  (c) *if $\varphi \in F$ and $\varphi \rightarrow \psi \in F$, then $\psi \in F$,*
  (d) *if $\vdash \varphi$, then $\varphi \in F$.*

The completeness of ALX means that:

(A) For arbitrary formula set Δ and arbitrary formula ϕ, $\Delta \models \phi \Rightarrow \Delta \vdash_{\mathrm{ALX}} \phi$.

It actually turns out to be easier to show the following statement:

(B) For arbitrary formula set Δ, Δ is consistent with ALX $\Leftrightarrow$ Δ has an ALX model.

We show that (A) and (B) are equivalent:

Proof.

(B) $\Rightarrow$ (A).

$\Delta \not\vdash_{\mathrm{ALX}} \phi$
$\Rightarrow \Delta \cup \{\neg\phi\}$ is consistent with ALX (Definition of consistency)
$\Rightarrow \Delta \cup \{\neg\phi\}$ has an ALX model (B)
$\Rightarrow \exists M \in Mod(M \models \Delta$ and $M \models \neg\phi)$ (Definition of $\models$)
$\Rightarrow \exists M \in Mod(M \models \Delta$ and $M \not\models \phi)$ (Truth condition)
$\Rightarrow$ It is not the case that
$\forall M \in Mod(M \models \Delta \Rightarrow M \models \phi)$ (Meta-reasoning)
$\Rightarrow \Delta \not\models \phi$ (Definition of $\models$).

(A) $\Rightarrow$ (B).

Δ has no ALX model
$\Rightarrow \neg(\exists M \in Mod(M \models \Delta))$ (Definition of $\models$)
$\Rightarrow \forall M \in Mod(M \not\models \Delta)$ (Meta-reasoning)
$\Rightarrow \forall M \in Mod(M \not\models \Delta$ or $M \models \bot)$ (Meta-reasoning)
$\Rightarrow \forall M \in Mod(M \models \Delta \Rightarrow M \models \bot)$ (Meta-reasoning)
$\Rightarrow \Delta \models \bot$ (Definition of $\models$)
$\Rightarrow \Delta \vdash_{\mathrm{ALX}} \bot$ (A)
$\Rightarrow \Delta$ is inconsistent with ALX (Definition of consistency) □

Assume that we can construct a canonical model M_c where the possible worlds are maximal consistent sets, then, in order to show the completeness, we have to show that for any formula ϕ,

(1) $\phi \in w \Leftrightarrow w \in [\![\phi]\!]_{M_c}$,
(2) M_c is an ALX model.

So our task is to construct a canonical model that is an ALX model. First, we need two lemmas (for the action and the update operators, respectively) that ensure the existence of certain maximal consistent sets required in the construction of the canonical model. Let W_c be the set of all maximal consistent sets built from the elements of *FML*.

Lemma 29 (Action Lemma).

$$\forall w \in W_c(\langle a\rangle\phi \in w \Rightarrow (\exists z \in W_c)(\phi \in z \textit{ and } (\forall \psi \in z)(\langle a\rangle\psi \in w))).$$

Proof. Suppose that $\langle a\rangle\phi \in w$ and let $F = \{\phi\} \cup \{\psi : \neg\langle a\rangle\neg\psi \in w\}$. Let $w^* = \{\psi : \neg\langle a\rangle\neg\psi \in w\}$.

We show first that (1) w^* and (2) F are consistent. We then show (3) that we can always extend F to an F' such that F' satisfies the condition of the lemma, i.e. $F' = z$.

(1) We claim that w^* is consistent. This is implied by:

(1.1) Assume that $\bot \in w^*$ we then show that $\bot \in w^* \rightarrow \textit{False}$.

$\bot \in w^*$
$\Rightarrow \neg\langle a\rangle\neg\bot \in w$ (Definition of w^*)
$\Rightarrow \neg\langle a\rangle\top \in w$ (Propositional logic)
$\Rightarrow \neg\langle a\rangle\top \wedge \langle a\rangle\phi \in w$ (Assumption)
$\Rightarrow \neg\langle a\rangle\top \wedge \langle a\rangle\top \in w$ (MONA)
$\Rightarrow$ *False* (Maximal consistency of w)

(1.2) We show that $\phi_1, \phi_2 \in w^* \Rightarrow (\phi_1 \wedge \phi_2) \in w^*$.

$\phi_1, \phi_2 \in w^*$
$\Rightarrow \neg\langle a\rangle\neg\phi_1 \in w$ and $\neg\langle a\rangle\neg\phi_2 \in w$ (Definition of w^*)
$\Rightarrow \neg(\langle a\rangle\neg\phi_1 \vee \langle a\rangle\neg\phi_2) \in w$ (Propositional logic)
$\Rightarrow \neg(\langle a\rangle(\neg\phi_1 \vee \neg\phi_2)) \in w$ (A2)
$\Rightarrow \neg(\langle a\rangle\neg(\phi_1 \wedge \phi_2)) \in w$ (Propositional logic)
$\Rightarrow \phi_1 \wedge \phi_2 \in w^*$ (Definition of w^*)

(1.3) For arbitrary ψ, we must show that $\psi \in w^*, \neg\psi \in w^* \Rightarrow$ *False*.

$\psi \in w^*, \neg\psi \in w^*$
$\Rightarrow (\psi \wedge \neg\psi) \in w^*$ (1.2)
$\Rightarrow \bot \in w^*$ (Propositional logic)
$\Rightarrow$ *False* (1.1)

We conclude that w^* is consistent.

(2) We claim that F is consistent. This is implied by: (2.1) $\phi \rightarrow \bot \Rightarrow$ *False* and (2.2) for any $\psi \in w^*, \phi \wedge \psi \rightarrow \bot \Rightarrow$ *False*.

(2.1) Assume that $\phi \rightarrow \bot$, then:

$\phi \rightarrow \bot$
$\Rightarrow \langle a\rangle\bot \in w$ ($\langle a\rangle\phi \in w$ and (MONA))
$\Rightarrow \bot \in w$ (A1)
$\Rightarrow$ *False* (Maximal consistency of w)

(2.2) For arbitrary $\psi \in w^*$, assume that $\phi \wedge \psi \rightarrow \bot$. We show that $\phi \wedge \psi \rightarrow \bot \Rightarrow$ *False*.

$\phi \wedge \psi \rightarrow \bot$
$\Rightarrow (\phi \rightarrow \neg\psi)$ (Propositional logic)
$\Rightarrow (\phi \rightarrow \neg\psi)$ and $\psi \in w^*$ and $\langle a\rangle\phi \in w$ (Assumption)
$\Rightarrow (\phi \rightarrow \neg\psi)$ and $\neg\langle a\rangle\neg\psi \in w$ and $\langle a\rangle\phi \in w$ (Definition of w^*)
$\Rightarrow \neg\langle a\rangle\neg\psi \in w$ and $\langle a\rangle\neg\psi \in w$ (MONA)
$\Rightarrow$ *False* (Maximal consistency of w)

We conclude that F is consistent.

We show now that any maximal extension F' of F satisfies the lemma. So, let F' be an arbitrary maximal consistent extension of F. We must show that: (3.1) F' exists, (3.2) $\phi \in F'$, (3.3) $\langle a\rangle\psi \notin w \Rightarrow \psi \notin F'$.

(3.1) Straightforward from Lindenbaum's Lemma.
(3.2) From the definition of F'.
(3.3) We have:

$$
\begin{array}{ll}
& \langle a\rangle\psi \notin w \\
\Rightarrow & \neg\langle a\rangle\psi \in w \quad \text{(Maximal consistency of } w\text{)} \\
\Rightarrow & \neg\langle a\rangle\neg\neg\psi \in w \quad \text{(Propositional logic)} \\
\Rightarrow & \neg\psi \in w^* \quad \text{(Definition of } w^*\text{)} \\
\Rightarrow & \neg\psi \in F \quad \text{(Definition of } F\text{)} \\
\Rightarrow & \neg\psi \in F' \quad \text{(Definition of } F'\text{)} \\
\Rightarrow & \psi \notin F' \quad \text{(Maximal consistency of } F'\text{).} \quad \square
\end{array}
$$

The next lemma parallels the Action Lemma for the update operator.

Lemma 30 (Update Lemma).

$$\forall w \in W_c(\phi \circ \chi \in w \Rightarrow (\exists z \in W_c)(\phi \in z \textit{ and } (\forall \psi \in z)(\psi \circ \chi \in w))).$$

Proof. Suppose that $\phi \circ \chi \in w$. Let

$$F = \{\phi\} \cup \{\psi : \neg(\neg\psi \circ \chi) \in w\}, \qquad w^\circ = \{\psi : \neg(\neg\psi \circ \chi) \in w\}.$$

The proof's geometry parallels the Action Lemma. We show first that (1) w° and (2) F are consistent. We then show (3) that we can always extend F to an F' such that F' satisfies the condition of the lemma, i.e. $F' = z$.

(1) We claim that w° is consistent.

(1.1) Assume that $\bot \in w^\circ$, then we can show that $\bot \in w^\circ \Rightarrow \textit{False}$.

$$
\begin{array}{ll}
& \bot \in w^\circ \\
\Rightarrow & \neg(\neg\bot \circ \chi) \in w \quad \text{(Definition of } w^\circ\text{)} \\
\Rightarrow & \neg(\top \circ \chi) \in w \quad \text{(Propositional logic)} \\
\Rightarrow & \neg(\top \circ \chi) \wedge (\phi \circ \chi) \in w \quad \text{(Assumption)} \\
\Rightarrow & \neg(\top \circ \chi) \wedge (\top \circ \chi) \in w \quad \text{(MONU)} \\
\Rightarrow & \textit{False} \quad \text{(Maximal consistency of } w\text{)}
\end{array}
$$

(1.2) We show that $\phi_1, \phi_2 \in w^\circ \Rightarrow (\phi_1 \wedge \phi_2) \in w^\circ$.

$$
\begin{array}{ll}
& \phi_1, \phi_2 \in w^\circ \\
\Rightarrow & \neg(\neg\phi_1 \circ \chi) \in w \text{ and } \neg(\neg\phi_2 \circ \chi) \in w \quad \text{(Definition of } w^\circ\text{)} \\
\Rightarrow & \neg(\neg\phi_1 \circ \chi \vee \neg\phi_2 \circ \chi) \in w \quad \text{(Maximal consistency of } w\text{)} \\
\Rightarrow & \neg((\neg\phi_1 \vee \neg\phi_2) \circ \chi) \in w \quad \text{(U4)} \\
\Rightarrow & \neg(\neg(\phi_1 \wedge \phi_2) \circ \chi) \in w \quad \text{(Propositional logic)} \\
\Rightarrow & \phi_1 \wedge \phi_2 \in w^\circ \quad \text{(Definition of } w^\circ\text{)}
\end{array}
$$

(1.3) For arbitrary ψ, we must show that $\psi \in w^\circ, \neg\psi \in w^\circ \Rightarrow \mathit{False}$.

$$
\begin{array}{ll}
& \psi \in w^\circ, \neg\psi \in w^\circ \\
\Rightarrow & (\psi \wedge \neg\psi) \in w^\circ \quad (1.2) \\
\Rightarrow & \bot \in w^\circ \quad \text{(Propositional logic)} \\
\Rightarrow & \mathit{False} \quad (1.1)
\end{array}
$$

We conclude that w° is consistent.

(2) We claim that F is consistent. This is implied by: (2.1) $\phi \rightarrow \bot \Rightarrow \mathit{False}$ and (2.2) For any $\psi \in w^\circ, \phi \wedge \psi \rightarrow \bot \Rightarrow \mathit{False}$.

(2.1) Assume that $\phi \rightarrow \bot$, then:

$$
\begin{array}{ll}
& \phi \rightarrow \bot \\
\Rightarrow & \bot \circ \chi \in w \quad (\phi \circ \chi \in w \text{ and (MONA)}) \\
\Rightarrow & \bot \in w \quad \text{(U3)} \\
\Rightarrow & \mathit{False} \quad \text{(Maximal consistency of } w\text{)}
\end{array}
$$

(2.2) For arbitrary $\psi \in w^\circ$, assume that $\phi \wedge \psi \rightarrow \bot$. We show that $\phi \wedge \psi \rightarrow \bot \Rightarrow \mathit{False}$.

$$
\begin{array}{lll}
& \phi \wedge \psi \rightarrow \bot & \\
\Rightarrow & (\phi \rightarrow \neg\psi) & \text{(Propositional logic)} \\
\Rightarrow & (\phi \rightarrow \neg\psi) \text{ and } \psi \in w^\circ \text{ and } \phi \circ \chi \in w & \text{(Assumption)} \\
\Rightarrow & (\phi \rightarrow \neg\psi) \text{ and } \neg\neg\psi \circ \chi \in w \text{ and } \phi \circ \chi \in w & \text{(Definition of } w^\circ\text{)} \\
\Rightarrow & (\neg\neg\psi \circ \chi) \in w \text{ and } (\neg\psi \circ \chi) \in w & \text{(MONU)} \\
\Rightarrow & \mathit{False} & \text{(Maximal consistency of } w\text{)}
\end{array}
$$

We conclude that F is consistent. We now show that any arbitrary maximal extension F' of F satisfies the lemma. So, let F' be an arbitrary maximal consistent extensions of F. We must show that: (3.1) F' exists, (3.2) $\phi \in F'$, (3.3) $\psi \circ \chi \notin w \Rightarrow \psi \notin F'$.

(3.1) Straightforward from Lindenbaum's Lemma.
(3.2) From the definition of F'.
(3.3) We show that as follows:

$$
\begin{array}{ll}
& \psi \circ \chi \notin w \\
\Rightarrow & \neg(\psi \circ \chi) \in w \quad \text{(Maximal consistency of } w\text{)} \\
\Rightarrow & \neg(\neg\neg\psi \circ \chi) \in w \quad \text{(Maximal consistency of } w\text{)} \\
\Rightarrow & \neg\psi \in w^\circ \quad \text{(Definition of } w^\circ\text{)} \\
\Rightarrow & \neg\psi \in F \quad \text{(Definition of } F\text{)} \\
\Rightarrow & \neg\psi \in F' \quad \text{(Definition of } F'\text{)} \\
\Rightarrow & \psi \notin F' \quad \text{(Maximal consistency of } F'\text{).} \quad \square
\end{array}
$$

Proposition 8 (Completeness of ALXS). *ALXS is complete for the class of ALX models.*

Proof. We construct a canonical model $M_c = \langle W_c, cw, R^a, \succ, V\rangle$ and show that:
(1) Truth Lemma: $\chi \in w \in W_c \Leftrightarrow w \in [\![\chi]\!]_{M_c}$.
(2) M_c is an ALX model.
Define $M_c = \langle W_c, cw, R^a, \succ, V\rangle$ as follows:

$$W_c = \{i : i \text{ is a maximal consistent set}\},$$
$$w \in cw(j, \{w' : \psi \in w'\}) \quad \text{iff} \quad \forall\rho(\rho \in j \Rightarrow \rho \circ \psi \in w),$$
$$\langle w, x\rangle \in R^a \quad \text{iff} \quad \forall\rho(\rho \in x \Rightarrow \langle a\rangle\rho \in w),$$
$$cw(w, \{w' : \phi \wedge \neg\psi \in w'\}) \succ cw(w, \{w' : \neg\phi \wedge \psi \in w'\}) \quad \text{iff} \quad \phi \boldsymbol{P} \psi \in w,$$
$$V(p_i) = \{w : p_i \in w\}.$$

We prove the Truth Lemma by induction on the complexity of χ.

(1.1) $\chi \equiv p_i$.

$p_i \in w$
$\Leftrightarrow w \in V(p_i)$ (Definition of V)
$\Leftrightarrow w \in [\![p_i]\!]_{M_c}$ (Truth condition)

(1.2) $\chi \equiv \neg\phi$.

$\neg\phi \in w$
$\Leftrightarrow \phi \notin w$ (Maximal consistency of w)
$\Leftrightarrow w \notin [\![\phi]\!]_{M_c}$ (Induction hypothesis)
$\Leftrightarrow w \in [\![\neg\phi]\!]_{M_c}$ (Truth condition)

(1.3) $\chi \equiv \phi \wedge \psi$.

$\phi \wedge \psi \in w$
$\Leftrightarrow \phi, \psi \in w$ (Maximal consistency of w)
$\Leftrightarrow w \in [\![\phi]\!]_{M_c}$ and $w \in [\![\psi]\!]_{M_c}$ (Induction hypothesis)
$\Leftrightarrow w \in [\![\phi \wedge \psi]\!]_{M_c}$ (Truth condition)

(1.4) $\chi \equiv \langle a\rangle\phi$.

$\langle a\rangle\phi \in w$
$\Rightarrow \exists z \in W_c(\phi \in z \text{ and } \forall\psi \in z(\langle a\rangle\psi \in w))$ (Action Lemma)
$\Rightarrow \exists z(\phi \in z \text{ and } R^a wz)$ (Definition of R^a)
$\Rightarrow \exists z(z \in [\![\phi]\!]_{M_c} \text{ and } R^a wz)$ (Induction hypothesis)
$\Rightarrow w \in [\![\langle a\rangle\phi]\!]_{M_c}$ (Truth condition)

$w \in [\![\langle a\rangle\phi]\!]_{M_c}$
$\Leftrightarrow \exists z \in W_c(R^a wz \text{ and } z \in [\![\phi]\!]_{M_c})$ (Truth condition)
$\Leftrightarrow \exists z \in W_c(R^a wz \text{ and } \phi \in z)$ (Induction hypothesis)
$\Rightarrow \langle a\rangle\phi \in w$ (Definition of R^a)

(1.5) $\chi \equiv \phi \circ \psi$.

$$
\begin{array}{ll}
& \phi \circ \psi \in w \\
\Rightarrow & \exists z(\phi \in z \text{ and } (\forall \rho \in z)((\rho \circ \psi) \in w)) \quad \text{(Update lemma)} \\
\Rightarrow & \exists z(\phi \in z \text{ and } w \in cw(z, \{w' : \psi \in w'\})) \quad \text{(Definition of } cw\text{)} \\
\Rightarrow & \exists z(z \in [\![\phi]\!]_{M_c} \text{ and } w \in cw(z, [\![\psi]\!]_{M_c})) \quad \text{(Induction hypothesis)} \\
\Rightarrow & w \in [\![\phi \circ \psi]\!]_{M_c} \quad \text{(Truth condition)}
\end{array}
$$

$$
\begin{array}{ll}
& w \in [\![\phi \circ \psi]\!]_{M_c} \\
\Leftrightarrow & \exists z(z \in [\![\phi]\!]_{M_c} \text{ and } w \in cw(z, [\![\psi]\!]_{M_c})) \quad \text{(Truth condition)} \\
\Leftrightarrow & \exists z(\phi \in z \text{ and } w \in cw(z, \{w' : \psi \in w'\})) \quad \text{(Induction hypothesis)} \\
\Rightarrow & \phi \circ \psi \in w \quad \text{(Definition of } cw\text{)}
\end{array}
$$

(1.6) $\chi \equiv \phi \boldsymbol{P} \psi$.

$$
\begin{array}{ll}
& \phi \boldsymbol{P} \psi \in w \\
\Leftrightarrow & cw(w, \{w' : \phi \wedge \neg\psi \in w'\}) \succ \\
& cw(w, \{w' : \psi \wedge \neg\phi \in w'\}) \quad \text{(Definition of } \succ\text{)} \\
\Leftrightarrow & cw(w, \{w' : \phi \in w'\} \cap \{w' : \neg\psi \in w'\}) \succ \\
& cw(w, \{w' : \psi \in w'\} \cap \{w' : \neg\phi \in w'\}) \quad \text{(Meta-reasoning)} \\
\Leftrightarrow & cw(w, [\![\phi]\!]_{M_c} \cap [\![\neg\psi]\!]_{M_c}) \succ \\
& cw(w, [\![\psi]\!]_{M_c} \cap [\![\neg\phi]\!]_{M_c}) \quad \text{(Induction hypothesis)} \\
\Leftrightarrow & cw(w, [\![\phi \wedge \neg\psi]\!]_{M_c}) \succ cw(w, [\![\psi \wedge \neg\phi]\!]_{M_c}) \quad \text{(Truth condition)} \\
\Leftrightarrow & w \in [\![\phi \boldsymbol{P} \psi]\!]_{M_c} \quad \text{(Truth condition)}
\end{array}
$$

This concludes the proof of the Truth Lemma. We now show that M_c is an ALX model. So, we have to show that cw satisfies (CS1), (CS2), and (CSC). Moreover, we have to show that $\succ$ satisfies the transitivity and normality conditions.

(CS1) $w \in cw(j, [\![\psi]\!]_{M_c}) \Rightarrow w \in [\![\psi]\!]_{M_c}$.

$$
\begin{array}{ll}
& w \in cw(j, [\![\psi]\!]_{M_c}) \\
\Leftrightarrow & \forall \rho(\rho \in j \Rightarrow \rho \circ \psi \in w) \quad \text{(Definition of } cw\text{)} \\
\Rightarrow & \exists \rho(\rho \in j \text{ and } \rho \circ \psi \in w) \quad (j \text{ is not an empty set}) \\
\Rightarrow & \psi \in w \quad \text{(U1)} \\
\Leftrightarrow & w \in [\![\psi]\!]_{M_c} \quad \text{(Truth Lemma)}
\end{array}
$$

(CS2) $j \in [\![\psi]\!]_{M_c} \Rightarrow cw(j, [\![\psi]\!]_{M_c}) = \{j\}$. We must show that: (a) $j \in [\![\psi]\!]_{M_c} \Rightarrow j \in cw(j, [\![\psi]\!]_{M_c})$, and (b) $j \in [\![\psi]\!]_{M_c}$ and $j' \in cw(j, [\![\psi]\!]_{M_c}) \Rightarrow j = j'$.

For (a), we have:

$$
\begin{array}{ll}
& j \in [\![\psi]\!]_{M_c} \\
\Leftrightarrow & \psi \in j \quad \text{(Truth Lemma)} \\
\Rightarrow & \forall \rho(\rho \in j \Rightarrow (\rho \wedge \psi) \in j) \quad \text{(Maximal consistency of } j\text{)} \\
\Rightarrow & \forall \rho(\rho \in j \Rightarrow (\rho \circ \psi) \in j) \quad \text{(U2)} \\
\Rightarrow & j \in cw(j, [\![\psi]\!]_{M_c}) \quad \text{(Definition of } cw\text{)}.
\end{array}
$$

For (b), suppose that $j \in [\![\psi]\!]_{M_c}$ and $j' \in cw(j, [\![\psi]\!]_{M_c})$, we first show that $j \subseteq j'$. Then by the maximal consistency of both j and j', we have $j = j'$. To show that $j \subseteq j'$, we proceed by reductio ad absurdum and show that $\rho \in j$ and $\rho \notin j' \Rightarrow$ *False* for arbitrary ρ.

$$
\begin{array}{ll}
& \rho \in j \text{ and } \rho \notin j' \\
\Leftrightarrow & \rho \in j \text{ and } \neg\rho \in j' \qquad \text{(Maximal consistency of } j'\text{)} \\
\Rightarrow & \rho \wedge \psi \in j \text{ and } \neg\rho \in j' \qquad (j \in [\![\psi]\!] \text{ and maximal consistency of } j) \\
\Rightarrow & ((\rho \wedge \psi) \circ \psi) \in j' \text{ and } \neg\rho \in j' \qquad (j' \in cw(j, [\![\psi]\!]_{M_c}) \text{ and definition of } cw) \\
\Rightarrow & ((\rho \wedge \psi) \circ \psi) \in j' \text{ and} \\
& \neg((\rho \wedge \psi) \circ \psi) \in j' \qquad \text{(U5)} \\
\Rightarrow & \textit{False} \qquad \text{(Maximal consistency of } j'\text{)}
\end{array}
$$

(CSC) $cw(w, [\![\phi]\!]_{M_c}) \cap [\![\psi]\!]_{M_c} \subseteq cw(w, [\![\phi \wedge \psi]\!]_{M_c})$. For any $j \in cw(w, [\![\phi]\!]_{M_c}) \cap [\![\psi]\!]_{M_c}$, we have to show that $j \in cw(w, [\![\phi \wedge \psi]\!]_{M_c})$. That is, for any ρ, if $\rho \in w$, then $\rho \circ (\phi \wedge \psi) \in j$ by the definition of cw.

For any ρ:

$$
\begin{array}{ll}
& \rho \in w \text{ and } j \in cw(w, [\![\phi]\!]_{M_c}) \cap [\![\psi]\!]_{M_c} \\
\Rightarrow & \rho \in w \text{ and } j \in cw(w, [\![\phi]\!]_{M_c}) \text{ and } \psi \in j \qquad \text{(Truth Lemma)} \\
\Rightarrow & \rho \circ \phi \in j \text{ and } \psi \in j \qquad \text{(Definition of } cw\text{)} \\
\Rightarrow & (\rho \circ \phi) \wedge \psi \in j \qquad \text{(Consistency of } w\text{)} \\
\Rightarrow & \rho \circ (\phi \wedge \psi) \in j \qquad \text{(U6)}.
\end{array}
$$

Therefore, $j \in cw(w, [\![\phi \wedge \psi]\!]_{M_c})$ by the definition of cw, so (CSC) holds.

(NORM) $(\emptyset \not\succ X)$. We must show that $\emptyset \succ X \Rightarrow$ **False**.

$$
\begin{array}{ll}
& \emptyset \succ X \\
\Rightarrow & \exists w \exists \phi \exists \psi (\phi \boldsymbol{P} \psi \in w \text{ and} \\
& \qquad cw(w, [\![\phi \wedge \neg\psi]\!]_{M_c}) = \emptyset \text{ and} \\
& \qquad cw(w, [\![\psi \wedge \neg\phi]\!]_{M_c}) = X \text{ and} \\
& \qquad cw(w, [\![\phi \wedge \neg\psi]\!]_{M_c}) \succ \\
& \qquad cw(w, [\![\psi \wedge \neg\phi]\!]_{M_c})) \qquad \text{(Definition of } \succ\text{)} \\
\Rightarrow & cw(w, [\![\bot]\!]_{M_c}) \succ cw(w, [\![\psi \wedge \neg\phi]\!]_{M_c}) \qquad (cw(w, [\![\bot]\!]_{M_c}) = \emptyset \text{ by (CS!)}) \\
\Rightarrow & cw(w, [\![\bot \wedge \neg(\psi \wedge \neg\phi)]\!]_{M_c}) \succ \\
& cw(w, [\![(\psi \wedge \neg\phi) \wedge \neg\bot]\!]_{M_c}) \qquad \text{(Meta-reasoning)} \\
\Rightarrow & \bot \boldsymbol{P} (\psi \wedge \neg\phi) \in w \qquad \text{(Definition of } cw\text{)} \\
\Rightarrow & \textbf{False} \qquad \text{(N)}
\end{array}
$$

(TRAN) $cw(w, X \cap \overline{Y}) \succ cw(w, Y \cap \overline{X})$ and $cw(w, Y \cap \overline{Z}) \succ cw(w, Z \cap \overline{Y}) \Rightarrow cw(w, X \cap \overline{Z}) \succ cw(w, Z \cap \overline{X})$.

$cw(w, X \cap \overline{Y}) \succ cw(w, Y \cap \overline{X})$ and
$cw(w, Y \cap \overline{Z}) \succ cw(w, Z \cap \overline{Y})$
$\Rightarrow \exists\phi\exists\psi\exists\chi(\phi \boldsymbol{P} \psi \in w$ and $\psi \boldsymbol{P} \chi \in w$ and
$[\![\phi]\!]_{M_c} = X$ and $[\![\psi]\!]_{M_c} = Y$ and $[\![\chi]\!]_{M_c} = Z)$ (Definition of $\succ$)
$\Rightarrow \exists\phi\exists\psi\exists\chi(\phi \boldsymbol{P} \chi \in w$ and $\psi \boldsymbol{P} \chi \in w$ and
$[\![\phi]\!]_{M_c} = X$ and $[\![\psi]\!]_{M_c} = Y$ and $[\![\chi]\!]_{M_c} = Z)$ (TR)
$\Rightarrow \exists\phi\exists\psi\exists$
$\chi(cw(w, [\![\phi \wedge \neg\chi]\!]_{M_c}) \succ \chi(cw(w, [\![\chi \wedge \neg\phi]\!]_{M_c})$ and
$[\![\phi]\!]_{M_c} = X$ and $[\![\psi]\!]_{M_c} = Y$ and $[\![\chi]\!]_{M_c} = Z)$ (Definition of $\succ$)
$\Rightarrow cw(w, X \cap \overline{Z}) \succ cw(w, Z \cap \overline{X})$ (Meta-reasoning)

This concludes the proof that M_c is an ALX model. □

A.5. The finite model property of ALX

Definition 31 (*Subformula set*). A formula set Φ_ρ is said to be the subformula set of ρ iff Φ_ρ satisfies the following conditions:

- $\rho \in \Phi_\rho$,
- $\neg\phi \in \Phi_\rho \Rightarrow \phi \in \Phi_\rho$,
- $\phi \wedge \psi \in \Phi_\rho \Rightarrow \phi, \psi \in \Phi_\rho$,
- $\langle a \rangle \phi \in \Phi_\rho \Rightarrow \phi \in \Phi_\rho$,
- $\phi \circ \psi \in \Phi_\rho \Rightarrow \phi, \psi \in \Phi_\rho$,
- $\phi \boldsymbol{P} \psi \in \Phi_\rho \Rightarrow \phi, \psi \in \Phi_\rho$.

Claim 32. *For any formula ρ, the subformula set of ρ is closed under subformulas.*

In ALX, the truth condition of $\phi \boldsymbol{P} \psi$ depends on the conjunction expansion principle $cw(w, [\![\phi \wedge \neg\psi]\!]_M) \succ cw(w, [\![\psi \wedge \neg\phi]\!]_M)$. Because of this, we shall need an extended subformula set to handle the problem. We define the extended subformula set accordingly.

Definition 33 (*Extended subformula set*). Let Φ be a formula set which is closed under subformulas. The extended subformula set Φ^{++} is defined as the Boolean closure of Φ. Let Φ^+ be the set of single representatives for each propositional equivalence class of the formulas in Φ^{++}. In particular, we have $\Phi \subseteq \Phi^+$.

It is easy to see that Φ^+ is finite, since Φ is finite. Moreover, for any $\phi, \psi \in \Phi^+$, there exist formulas $\chi_1, chi_2 \in \Phi^+$ such that $\vdash (\chi_1 \leftrightarrow (\phi \wedge \neg\psi))$ and $\vdash (\chi_2 \leftrightarrow (\psi \wedge \neg\phi))$. In particular, we have $\bot \in \Phi^+$.

Definition 34 (*Equivalence relation on possible worlds*). Let M be an ALX model $\langle W, \succ, cw, R^a, V \rangle$ and Φ be a formula set which is closed under subformulas. For any $w, w' \in W$, we define

$w \approx w'$ with respect to M and Φ^+ iff $\forall \rho \in \Phi^+ (M, w \Vdash \rho \Leftrightarrow M, w' \Vdash \rho)$.

Definition 35 (*Equivalence class*). Let $\approx$ be an equivalence relation on possible worlds with respect to M and Φ^+ and w be a possible world, we define

$$[w] \stackrel{\text{def}}{\Longleftrightarrow} \{w' \in W : \ w \approx w'\}.$$

Definition 36 (*Filtration*). A filtration of $M = \langle W, \succ, cw, R^a, V\rangle$ through Φ^+ is any model $M^* = \langle W^*, \succ^*, cw^*, R^{a*}, V^*\rangle$ which satisfies the following conditions:

(1) W^* is a subset of W which consists of exactly one world from each equivalence class.

(2) $R^{a*}, cw^*, \succ^*$ satisfy the following suitability conditions:

(2.1) $\forall w, w' \in W^*((\exists u \in W)(R^a wu$ and $w' \approx u) \Rightarrow R^{a*}ww')$,

(2.2) $\forall w, w' \in W^*$
$(R^{a*}ww' \Rightarrow (\forall\langle a\rangle\phi \in \Phi^+)(M, w' \Vdash \phi \Rightarrow M, w \Vdash \langle a\rangle\phi))$,

(2.3) $\forall w, w' \in W^*$
$((\forall\psi \in \Phi^+)((\exists u \in W)\ (w \in cw(u, [\![\psi]\!]_M)$ and $w' \approx u) \Rightarrow$
$w \in cw^*(w', [\![\psi]\!]_{M^*})))$,

(2.4) $\forall w, w' \in W^*$
$((\forall\psi \in \Phi^+)$
$(w \in cw^*(w', [\![\psi]\!]_{M^*}) \Rightarrow$
$(\forall\phi \in \Phi^+)$
$((M, w' \Vdash \phi \wedge \psi \Rightarrow M, w \Vdash (\phi \wedge \psi) \circ \psi)$ and
$(M, w' \Vdash \phi \wedge \neg\psi \Rightarrow M, w \Vdash (\phi \wedge \neg\psi) \circ \psi)$ and
$(M, w' \Vdash (\neg\phi \wedge \psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \psi) \circ \psi)$ and
$(M, w' \Vdash (\neg\phi \wedge \neg\psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \neg\psi) \circ \psi))))$,

(2.5) $\forall w, \in W^*(\forall\phi, \psi \in \Phi^+)$
$(cw^*(w, [\![\phi \wedge \neg\psi]\!]_{M^*}) \succ^* cw^*(w, [\![\neg\phi \wedge \psi]\!]_{M^*}) \Leftrightarrow M, w \Vdash (\phi \boldsymbol{P} \psi))$.

(3) $V^*(p_i) = V(p_i)$ for any $p_i \in \Phi^+$.

Theorem 37 (Filtration Theorem). *Let $M = \langle W, \succ, cw, R^a, V\rangle$ be any ALX model, Φ be any formula set which is closed under subformulas and $M^* = \langle W^*, \succ^*, cw^*, R^{a*}, V^*\rangle$ be any filtration of M through Φ^+, then for any $\chi \in \Phi^+$ and any $w \in W^*$ $(M, w \Vdash \chi \Leftrightarrow M^*, w \Vdash \chi)$.*

Proof. We prove the theorem by induction on the complexity of Φ^+. For any $\chi \in \Phi$:

(1) $\chi \equiv p_i$.

$M, w \Vdash p_i$
$\Leftrightarrow w \in V(p_i)$ (Truth condition)
$\Leftrightarrow w \in V^*(p_i)$ (Definition of V^*)
$\Leftrightarrow M^*, w \Vdash p_i$ (Truth condition)

(2) $\chi \equiv \neg\phi$. We know that $\phi \in \Phi^+$,

$M, w \Vdash \neg\phi$
$\Leftrightarrow M, w \nVdash \phi$ (Truth condition)
$\Leftrightarrow M^*, w \nVdash \phi$ (Induction hypothesis)
$\Leftrightarrow M^*, w \Vdash \neg\phi$ (Truth condition)

(3) $\chi \equiv \phi \wedge \psi$. We know that $\phi, \psi \in \Phi^+$,

$M, w \Vdash \phi \wedge \psi$
$\Leftrightarrow M, w \Vdash \phi$ and $M, w \Vdash \psi$ (Truth condition)
$\Leftrightarrow M^*, w \Vdash \phi$ and $M^*, w \Vdash \psi$ (Induction hypothesis)
$\Leftrightarrow M^*, w \Vdash \phi \wedge \psi$ (Truth condition)

(4) $\chi \equiv \langle a \rangle \phi$. We know that $\phi \in \Phi^+$,

$M, w \Vdash \langle a \rangle \phi$
$\Rightarrow \exists u \in W(R^a wu$ and $M, u \Vdash \phi)$ (Truth condition)
$\Rightarrow \exists w' \in W^*(R^a wu$ and $w' \approx u$ and $M, u \Vdash \phi)$ (Definition of W^*)
$\Rightarrow \exists w' \in W^*(R^a wu$ and $w' \approx u$ and $M, w' \Vdash \phi)$ (Definition of $\approx$)
$\Rightarrow \exists w' \in W^*(R^{a*} ww'$ and $M, w' \Vdash \phi)$ (2.1)
$\Rightarrow \exists w' \in W^*(R^{a*} ww'$ and $M^*, w' \Vdash \phi)$ (Induction hypothesis)
$\Rightarrow M^*, w \Vdash \langle a \rangle \phi$ (Truth condition)

$M^*, w \Vdash \langle a \rangle \phi$
$\Leftrightarrow \exists w' \in W^*(R^{a*} ww'$ and $w' \in [\![\phi]\!]_{M^*})$ (Truth condition)
$\Leftrightarrow \exists w' \in W^*(R^{a*} ww'$ and $w' \in [\![\phi]\!]_M)$ (Induction hypothesis)
$\Rightarrow M, w \Vdash \langle a \rangle \phi$ (2.2)

(5) $\chi \equiv \phi \circ \psi$. We know that $\phi, \psi \in \Phi^+$,

$M, w \Vdash \phi \circ \psi$
$\Rightarrow \exists u(M, u \Vdash \phi$ and $w \in cw(u, [\![\psi]\!]_M))$ (Truth condition)
$\Rightarrow \exists w' \in W^*(M, u \Vdash \phi$ and $w' \approx u$ and
$w \in cw(u, [\![\psi]\!]_M))$ (Definition of W^*)
$\Rightarrow \exists w' \in W^*(M, w' \Vdash \phi$ and $w \in cw(u, [\![\psi]\!]_M))$ (Definition of $\approx$)
$\Rightarrow \exists w' \in W^*(M, w' \Vdash \phi$ and $w \in cw^*(w', [\![\psi]\!]_{M^*}))$ (2.3)
$\Rightarrow \exists w' \in W^*(M^*, w' \Vdash \phi$ and $w \in cw^*(w', [\![\psi]\!]_{M^*}))$ (Induction hypothesis)
$\Rightarrow M^*, w \Vdash \phi \circ \psi$ (Truth condition)

$M^*, w \Vdash \phi \circ \psi$
$\Leftrightarrow \exists w' \in W^*(M^*, w' \Vdash \phi$ and $w \in cw^*(w', [\![\psi]\!]_{M^*}))$ (Truth condition)
$\Leftrightarrow \exists w' \in W^*(M, w' \Vdash \phi$ and $w \in cw^*(w', [\![\psi]\!]_{M^*}))$ (Induction hypothesis)

Case 1: $M, w' \Vdash \psi$.

$$
\begin{aligned}
& M, w' \Vdash \psi \text{ and } M, w' \Vdash \phi \text{ and } w \in cw^*(w', [\![\psi]\!]_{M^*})) \\
\Rightarrow\ & M, w' \Vdash (\phi \wedge \psi) \text{ and } w \in cw^*(w', [\![\psi]\!]_{M^*})) && \text{(Truth condition)} \\
\Rightarrow\ & M, w \Vdash (\phi \wedge \psi) \circ \psi && (2.4) \\
\Rightarrow\ & M, w \Vdash \phi \circ \psi && \text{(MONU)}
\end{aligned}
$$

Case 2: $M, w' \Vdash \neg\psi$.

$$
\begin{aligned}
& M, w' \Vdash \neg\psi \text{ and } M, w' \Vdash \phi \text{ and } w \in cw^*(w', [\![\psi]\!]_{M^*})) \\
\Rightarrow\ & M, w' \Vdash (\phi \wedge (\neg\psi)) \text{ and } w \in cw^*(w', [\![\psi]\!]_{M^*})) && \text{(Truth condition)} \\
\Rightarrow\ & M, w \Vdash (\phi \wedge (\neg\psi)) \circ \psi && (2.4) \\
\Rightarrow\ & M, w \Vdash \phi \circ \psi && \text{(MONU)}
\end{aligned}
$$

(6) $\chi \equiv \phi \boldsymbol{P} \psi$. We know that $\phi, \psi \in \Phi^+$,

$$
\begin{aligned}
& M, w \Vdash \phi \boldsymbol{P} \psi \\
\Leftrightarrow\ & cw^*(w, [\![\phi \wedge \neg\psi]\!]_M) \succ^* cw^*(w, [\![\psi \wedge \neg\phi]\!]_{M^*}) && (2.5) \\
\Leftrightarrow\ & M^*, w \Vdash \phi \boldsymbol{P} \psi && \text{(Truth condition)}
\end{aligned}
$$

For any χ such that $\chi \in \Phi^+$ but $\chi \notin \Phi$, we know the following facts:

(7) $\chi \equiv \neg\phi$ and $\phi \in \Phi$. So $\phi \in \Phi^+$.

$$
\begin{aligned}
& M, w \Vdash \neg\phi \\
\Leftrightarrow\ & M, w \nVdash \phi && \text{(Truth condition)} \\
\Leftrightarrow\ & M^*, w \nVdash \phi && \text{(Induction hypothesis)} \\
\Leftrightarrow\ & M^*, w \Vdash \neg\phi && \text{(Truth condition)}
\end{aligned}
$$

(8) $\chi \equiv \phi \wedge \psi$ and $\phi, \psi \in \Phi$.

$$
\begin{aligned}
& M, w \Vdash \phi \wedge \psi \\
\Leftrightarrow\ & M, w \Vdash \phi \text{ and } M, w \Vdash \psi && \text{(Truth condition)} \\
\Leftrightarrow\ & M^*, w \Vdash \phi \text{ and } M^*, w \Vdash \psi && \text{(Induction hypothesis)} \\
\Leftrightarrow\ & M^*, w \Vdash \phi \wedge \psi && \text{(Truth condition)}
\end{aligned}
$$

Therefore, for any $\chi \in \Phi^+$, we have $M, w \Vdash \chi \leftrightarrow M^*, w \Vdash \chi$. □

Corollary 38 (Filtration Corollary). *Let* $M = \langle W, \succ, cw, R^a, V \rangle$ *be any ALX model,* Φ *be any formula set which is closed under subformulas and* $M^* = \langle W^*, \succ^*, cw^*, R^{a*}, V^* \rangle$ *be any filtration of* M *through* Φ^+*, then for any* $\phi, \psi \in \Phi^+$ *and* $w \in W^*$:

(a) $M, w \Vdash \neg\phi \Leftrightarrow M^*, w \Vdash \neg\phi$,
(b) $M, w \Vdash \phi \wedge \psi \Leftrightarrow M^*, w \Vdash \phi \wedge \psi$,
(c) $M, w \Vdash \phi \wedge \neg\psi \Leftrightarrow M^*, w \Vdash \phi \wedge \neg\psi$.

Proof. Straightforward. □

Theorem 39 (Invalidity Theorem). *Suppose that a formula χ is invalid in a model M, then χ is invalid in every filtration of M through Φ_χ^+.*

Proof. Since χ is invalid in a model $M = \langle W, \succ, cw, R^a, V\rangle$, there is some $w \in W$ such that $M, w \Vdash \neg\chi$. Suppose that $M^* = \langle W^*, cw^*, \succ^*, R^{a*}, V^*\rangle$ is a filtration of M through Φ_χ^+. By the definition of W^*, there is some $w^* \in W^*$ such that $w \approx w^*$ with respect to M and Φ_χ^+. Obviously, $\chi \in \Phi_\chi^+$, therefore $M, w^* \Vdash \neg\chi$. By Corollary 38(a), $M^*, w^* \Vdash \neg\chi$ and so χ is invalid in M^*. □

Theorem 10. *ALX has the finite model property.*

Proof. For arbitrary χ, suppose that $\nvdash_{\mathrm{ALX}} \chi$, then there exists a model $M = \langle W, \succ, cw, R^a, V\rangle$ and a world $w \in W$ such that $M, w \nVdash \chi$. Let Φ_χ be the subformula set of χ. We know that Φ_χ is finite. Moreover, Φ_χ^+ also is finite, by the definition of Φ_χ^+.

Now, we construct a filtration $M^* = \langle W^*, \succ^*, cw^*, R^{a*}, V^*\rangle$ of M through Φ_χ^+ as follows:

(1) For W^*, we first construct the equivalence class [ ] on W as:

$$[w] \overset{\text{def}}{\Longleftrightarrow} \{w' : \forall\rho \in \Phi_\chi^+ \forall w' \in W(M, w \Vdash \rho \Leftrightarrow M, w' \Vdash \rho)\}.$$

From each class $[w]$, we select exactly one world $w' \in [w]$ to represent this class. Now let W^* be the set of all representing worlds.

From the definition of the equivalence class [ ], we know that for any class $[w1]$ and any class $[w2]$, if $[w1] \neq [w2]$, then there exists $\rho \in \Phi_\chi^+$ such that either $M, w1 \Vdash \rho$ and $M, w2 \nVdash \rho$, or $M, w1 \nVdash \rho$ and $M, w2 \Vdash \rho$. Because Φ_χ^+ is finite, there are only finitely many formulas ρ by which we can distinguish two different classes. Therefore, there are only finitely many equivalence classes, namely, at most $2^{\mathrm{Card}(\Phi_\chi^+)}$. So W^* is finite.

(2) For V^*, we define

$$V^*(p_i) \overset{\text{def}}{\Longleftrightarrow} V(p_i) \quad \text{if } p_i \in \Phi_\chi^+.$$

(3) For R^{a*}, we define that, for any $w, w' \in W^*$,

$$\langle w, w'\rangle \in R^{a*} \quad \text{iff} \quad (\forall\langle a\rangle\phi \in \Phi_\chi^+)(M, w' \Vdash \phi \Rightarrow M, w \Vdash \langle a\rangle\phi).$$

(4) For cw^*, we define that, for any $w, w' \in W^*$ and any $\psi \in \Phi_\chi^+$,

$$\begin{aligned} &w \in cw^*(w', [\![\psi]\!]_{M^*}) \quad \text{iff} \\ &((\forall\phi \in \Phi_\chi^+)(\ (M, w' \Vdash \phi \wedge \psi \Rightarrow M, w \Vdash M, w \Vdash (\phi \wedge \psi) \circ \psi) \text{ and} \\ &\qquad\qquad (M, w' \Vdash \phi \wedge \neg\psi \Rightarrow M, w \Vdash (\phi \wedge \neg\psi) \circ \psi) \text{ and} \\ &\qquad\qquad (M, w' \Vdash (\neg\phi \wedge \psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \psi) \circ \psi) \text{ and} \\ &\qquad\qquad (M, w' \Vdash (\neg\phi \wedge \neg\psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \neg\psi) \circ \psi))). \end{aligned}$$

(5) For $\succ^*$, we define that, for any $w \in W^*$ and any $\phi, \psi \in \Phi_\chi^+$,

$$cw^*(w, [\![\phi \wedge \neg\psi]\!]_{M^*}) \succ^* cw^*(w, [\![\neg\phi \wedge \psi]\!]_{M^*}) \quad \text{iff} \quad M, w \Vdash \phi \boldsymbol{P} \psi.$$

Now, we have to show that M^* satisfies the conditions of a filtration of M through Φ_χ^+. From the construction of W^*, we know that W^* is a subset of W. Moreover, W^* consists of exactly one world from each equivalence class with respect to M and Φ_χ^+. Therefore, the condition for W^* is satisfied.

From the above definition of V^*, R^{a*}, cw^*, $\succ^*$, the suitability conditions (2.2), (2.4), (2.5) of Definition 36 and the condition for V^* are obviously satisfied.

To show that (2.1) is satisfied, we have to show that

$$\forall w, w' \in W^*(\exists u \in W)(w' \approx u \text{ and } R^a wu)) \Rightarrow R^{a*} ww').$$

Suppose that $\exists u \in W(w' \approx u$ and $R^a wu)$ and for any $\langle a\rangle\phi \in \Phi_\chi^+$:

$$\begin{aligned}
&\langle a\rangle\phi \in \Phi_\chi^+ \text{ and } M, w' \Vdash \phi \text{ and } R^a wu \\
\Rightarrow\ &\langle a\rangle\phi \in \Phi_\chi^+ \text{ and } M, u \Vdash \phi \text{ and } R^a wu && \text{(Definition of } \approx) \\
\Rightarrow\ &\langle a\rangle\phi \in \Phi_\chi^+ \text{ and } M, w \Vdash \langle a\rangle\phi && \text{(Truth condition)}
\end{aligned}$$

Therefore, according to the definition of R^{a*}, we have $R^{a*}ww'$, so (2.1) holds.

To show that (2.3) is satisfied, we have to show that

$$\begin{aligned}
&(\forall w, w' \in W^*)(\forall \psi \in \Phi_\chi^+) \\
&\quad((\exists u \in W)(w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M)) \Rightarrow w \in cw^*(w', [\![\psi]\!]_{M^*})).
\end{aligned}$$

For any $w, w' \in W^*$ and any $\psi \in \Phi_\chi^+$, suppose that

$$(\exists u \in W)(w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M))$$

and for any $\phi \in \Phi_\chi^+$:

(1) Assume that $M, w' \Vdash (\phi \wedge \psi)$, then:

$$\begin{aligned}
&M, w' \Vdash (\phi \wedge \psi) \text{ and } w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M) \\
\Rightarrow\ &M, u \Vdash (\phi \wedge \psi) \text{ and } w \in cw(u, [\![\psi]\!]_M) && \text{(Definition of } \approx) \\
\Rightarrow\ &M, w \Vdash (\phi \wedge \psi) \circ \psi) && \text{(Truth condition)}
\end{aligned}$$

(2) Assume that $M, w' \Vdash (\phi \wedge \neg\psi)$, then:

$$\begin{aligned}
&M, w' \Vdash (\phi \wedge \neg\psi) \text{ and } w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M) \\
\Rightarrow\ &M, u \Vdash (\phi \wedge \neg\psi) \text{ and } w \in cw(u, [\![\psi]\!]_M) && \text{(Definition of } \approx) \\
\Rightarrow\ &M, w \Vdash (\phi \wedge \neg\psi) \circ \psi) && \text{(Truth condition)}
\end{aligned}$$

(3) Assume that $M, w' \Vdash (\neg\phi \wedge \psi)$, then:

$$\begin{aligned}
&M, w' \Vdash (\neg\phi \wedge \psi) \text{ and } w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M) \\
\Rightarrow\ &M, u \Vdash (\neg\phi \wedge \psi) \text{ and } w \in cw(u, [\![\psi]\!]_M) && \text{(Definition of } \approx) \\
\Rightarrow\ &M, w \Vdash (\neg\phi \wedge \psi) \circ \psi) && \text{(Truth condition)}
\end{aligned}$$

(4) Assume that $M, w' \Vdash (\neg\phi \wedge \neg\psi)$, then:

$$
\begin{aligned}
& M, w' \Vdash (\neg\phi \wedge \neg\psi) \text{ and } w' \approx u \text{ and } w \in cw(u, [\![\psi]\!]_M) \\
\Rightarrow\ & M, u \Vdash (\neg\phi \wedge \neg\psi) \text{ and } w \in cw(u, [\![\psi]\!]_M) && \text{(Definition of } \approx) \\
\Rightarrow\ & M, w \Vdash (\neg\phi \wedge \neg\psi) \circ \psi) && \text{(Truth condition)}
\end{aligned}
$$

Therefore, according to the definition of cw^* above, we have $w \in cw^*(w', [\![\psi]\!]_{M^*})$. So (2.3) holds.

We know now that M^* is indeed a filtration of M through Φ_χ^+. Moreover, we know that M^* is a finite model. By the above theorem, we know that there exists a $w \in W^*$ such that $M^*, w \nVdash \chi$. Therefore, condition (1) of the finite model property (Definition 9) is satisfied.

In order to show that condition (2) of the finite model property is also satisfied, we have to show that M^* is an ALX model. That is to say, we have to show that cw^* satisfies (CS1)–(CSC) and $\succ^*$ satisfies the transitivity and the normality.

For any $w, w' \in W^*$ and any $\psi \in \Phi_\chi^+$,

(CS1) $w \in cw^*(w', [\![\psi]\!]_{M^*}) \Rightarrow w \in [\![\psi]\!]_{M^*}$.

$$
\begin{aligned}
& w \in cw^*(w', [\![\psi]\!]_{M^*}) \\
\Leftrightarrow\ & (\forall \phi \in \Phi_\chi^+) \\
& \quad ((M, w' \Vdash (\phi \wedge \psi) \Rightarrow M, w \Vdash (\phi \wedge \psi) \circ \psi) \text{ and} \\
& \quad (M, w' \Vdash (\phi \wedge \neg\psi) \Rightarrow M, w \Vdash (\phi \wedge \neg\psi) \circ \psi) \text{ and} \\
& \quad (M, w' \Vdash (\neg\phi \wedge \psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \psi) \circ \psi) \text{ and} \\
& \quad (M, w' \Vdash (\neg\phi \wedge \neg\psi) \Rightarrow M, w \Vdash (\neg\phi \wedge \neg\psi) \circ \psi)) && \text{(Definition of } cw^*)
\end{aligned}
$$

Case 1: $M, w' \Vdash (\phi \wedge \psi)$.

$$
\begin{aligned}
& M, w' \Vdash (\phi \wedge \psi) \\
\Rightarrow\ & M, w \Vdash (\phi \wedge \psi) \circ \psi && \text{(Definition of } cw^*) \\
\Rightarrow\ & M, w \Vdash \psi && (M \text{ is an ALX model and (U1)}) \\
\Rightarrow\ & M^*, w \Vdash \psi && \text{(Filtration Theorem)} \\
\Rightarrow\ & w \in [\![\psi]\!]_{M^*} && \text{(Definition of } [\![\,]\!]_{M^*})
\end{aligned}
$$

The other cases $(\phi \wedge \neg\psi, \neg\phi \wedge \psi, \neg\phi \wedge \neg\psi)$ are proved similarly. Therefore, (CS1) is satisfied.

(CS2) $w \in [\![\psi]\!]_{M^*} \Rightarrow cw^*(w, [\![\psi]\!]_{M^*}) = \{w\}$. We must show that:
(a) $w \in [\![\psi]\!]_{M^*} \Rightarrow w \in cw^*(w, [\![\psi]\!]_{M^*})$,
(b) $w \in [\![\psi]\!]_{M^*}$ and $w' \in cw^*(w, [\![\psi]\!]_{M^*}) \Rightarrow w \equiv w'$,
where $w \equiv w'$ means that w and w' represent the same equivalence class with respect to M^* and Φ_χ^+, namely,

$$\forall \rho \in \Phi_\chi^+ (M^*, w \Vdash \rho \Leftrightarrow M^*, w \Vdash \rho).$$

For (a), we will show that $w \in [\![\psi]\!]_{M^*}$ and $w \notin cw^*(w, [\![\psi]\!]_{M^*}) \Rightarrow \mathit{False}$.

$w \in [\![\psi]\!]_{M^*}$ and $w \notin cw^*(w, [\![\psi]\!]_{M^*})$

$\Rightarrow$ $w \in [\![\psi]\!]_M$ and $w \notin cw^*(w, [\![\psi]\!]_{M^*})$ (Filtration Theorem)

$\Rightarrow$ $w \in [\![\psi]\!]_M$ and $(\exists \phi \in \Phi_\chi^+)$
$((M, w \,|\!\!\sim (\phi \wedge \psi)$ and $M, w \,|\!\!\not\sim (\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\phi \wedge \neg\psi)$ and
$M, w \,|\!\!\not\sim (\phi \wedge \neg\psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\neg\phi \wedge \psi)$ and
$M, w \,|\!\!\not\sim (\neg\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\neg\phi \wedge \neg\psi)$ and
$M, w \,|\!\!\not\sim (\neg\phi \wedge \neg\psi) \circ \psi)$ (Definition of cw^*)

$\Rightarrow$ $w \in [\![\psi]\!]_M$ and $(\exists \phi \in \Phi_\chi^+)$
$((M, w \,|\!\!\sim (\phi \wedge \psi)$ and $M, w \,|\!\!\not\sim (\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\neg\psi)$ and $M, w \,|\!\!\not\sim (\phi \wedge \neg\psi) \circ \psi)$
or $(M, w \,|\!\!\sim \neg\phi \wedge \psi$ and
$M, w \,|\!\!\not\sim (\neg\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\neg\psi)$ and
$M, w \,|\!\!\not\sim (\neg\phi \wedge \neg\psi) \circ \psi)$ (MONU)

$\Rightarrow$ $w \in [\![\psi]\!]_M$ and $(\exists \phi \in \Phi_\chi^+)$
$((M, w \,|\!\!\sim (\phi \wedge \psi)$ and $M, w \,|\!\!\not\sim (\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim (\neg\phi \wedge \psi)$ and
$M, w \,|\!\!\not\sim (\neg\phi \wedge \psi) \circ \psi)$ (Meta-reasoning)

$\Rightarrow$ $\exists \phi \in \Phi_\chi^+$
$((M, w \,|\!\!\sim ((\phi \wedge \psi) \wedge \psi)$ and
$M, w \,|\!\!\not\sim (\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim ((\neg\phi \wedge \psi) \wedge \psi)$ and ($M, w \,|\!\!\sim \psi$,
$M, w \,|\!\!\not\sim (\neg\phi \wedge \psi) \circ \psi)$ and Truth condition)

$\Rightarrow$ $\exists \phi \in \Phi_\chi^+$
$((M, w \,|\!\!\sim ((\phi \wedge \psi) \circ \psi)$ and
$M, w \,|\!\!\not\sim (\phi \wedge \psi) \circ \psi)$
or $(M, w \,|\!\!\sim ((\neg\phi \wedge \psi) \circ \psi)$ and (M is an ALX model,
$M, w \,|\!\!\not\sim (\neg\phi \wedge \psi) \circ \psi)$ and (U2))

$\Rightarrow$ *False*

For (b), suppose that $w \in [\![\psi]\!]_{M^*}$ and $w' \in cw^*(w, [\![\psi]\!]_{M^*})$, for any $\phi \in \Phi_\chi^+$, we have to show that

$$M^*, w \,|\!\!\sim \rho \Leftrightarrow M^*, w' \,|\!\!\sim \rho.$$

($\Rightarrow$) We show that $M^*, w \Vdash \rho$ and $M^*, w' \nVdash \rho \Rightarrow$ *False*.

$M^*, w \Vdash \rho$ and $M^*, w' \nVdash \rho$
$\Rightarrow$ $M, w \Vdash \rho$ and $M^*, w' \nVdash \rho$ (Filtration Theorem)
$\Rightarrow$ $M, w \Vdash \rho$ and $M^*, w' \Vdash \neg\rho$ (Truth condition)
$\Rightarrow$ $M, w \Vdash \rho$ and $M, w' \Vdash \neg\rho$ (Filtration Corollary (a))
$\Rightarrow$ $M, w \Vdash \rho$ and $M, w' \Vdash \neg((\rho \wedge \psi) \circ \psi)$ (M is an ALX model and (U5))
$\Rightarrow$ $M, w \Vdash (\rho \wedge \psi)$ and
$M, w' \Vdash \neg((\rho \wedge \psi) \circ \psi)$ ($w \in [\![\psi]\!]_{M^*}$)
$\Rightarrow$ $M, w' \Vdash (\rho \wedge \psi) \circ \psi$ and
$M, w' \Vdash \neg((\rho \wedge \psi) \circ \psi)$ (Definition of cw^*)
$\Rightarrow$ *False*

($\Leftarrow$) We show that $M^*, w' \Vdash \rho$ and $M^*, w \nVdash \rho \Rightarrow$ *False*.

$M^*, w' \Vdash \rho$ and $M^*, w \nVdash \rho$
$\Rightarrow$ $M, w' \Vdash \rho$ and $M^*, w \nVdash \rho$ (Filtration Theorem)
$\Rightarrow$ $M, w' \Vdash \rho$ and $M^*, w \Vdash \neg\rho$ (Truth condition)
$\Rightarrow$ $M, w' \Vdash \rho$ and $M, w \Vdash \neg\rho$ (Filtration Corollary (a))
$\Rightarrow$ $M, w' \Vdash \rho$ and $M^*, w \Vdash \psi$ and $M, w \Vdash \neg\rho$ ($w \in [\![\psi]\!]_{M^*}$)
$\Rightarrow$ $M, w' \Vdash \rho$ and $M, w \Vdash \psi$ and $M, w \Vdash \neg\rho$ (Filtration Theorem)
$\Rightarrow$ $M, w' \Vdash \rho$ and $M, w \Vdash (\neg\rho \wedge \psi)$ (Truth condition)
$\Rightarrow$ $M, w' \Vdash \rho$ and $M, w' \Vdash (\neg\rho \wedge \psi) \circ \psi$ (Definition of cw^*)
$\Rightarrow$ $M, w' \Vdash \neg(\neg\rho)$ and $M, w' \Vdash (\neg\rho \wedge \psi) \circ \psi$ (Meta-reasoning)
$\Rightarrow$ $M, w' \Vdash \neg(((\neg\rho) \wedge \psi) \circ \psi)$ and (M is an ALX model
$M, w' \Vdash (\neg\rho \wedge \psi) \circ \psi$ and (U5))
$\Rightarrow$ *False*

(CSC) $j \in cw^*(w, [\![\phi]\!]_{M^*}) \cap [\![\psi]\!]_{M^*} \Rightarrow j \in cw^*(w, [\![\phi \wedge \psi]\!]_{M^*})$.

$j \in cw^*(w, [\![\phi]\!]_{M^*}) \cap [\![\psi]\!]_{M^*}$
$\Leftrightarrow$ $(\forall \rho \in \Phi_\chi^+)$
$((M, w \Vdash (\rho \wedge \phi) \Rightarrow M, j \Vdash (\rho \wedge \phi) \circ \phi)$ and
$(M, w \Vdash (\rho \wedge \neg\phi) \Rightarrow M, j \Vdash (\rho \wedge \neg\phi) \circ \phi)$ and
$(M, w \Vdash (\neg\rho \wedge \phi) \Rightarrow M, j \Vdash (\neg\rho \wedge \phi) \circ \phi)$ and
$(M, w \Vdash (\neg\rho \wedge \neg\phi) \Rightarrow$
$M, j \Vdash (\neg\rho \wedge \neg\phi) \circ \phi)$ and
$M^*, j \Vdash \psi)$ (Definition of cw^*)
$\Rightarrow$ $M, j \Vdash \psi$ (Filtration Lemma)

For any $\rho \in \Phi_\chi^+$:
Case 1: $M, w \Vdash (\rho \wedge (\phi \wedge \psi))$.

$$
\begin{aligned}
& M, w \Vdash (\rho \wedge (\phi \wedge \psi)) \\
\Rightarrow\ & M, w \Vdash (\rho \wedge \phi) && \text{(Meta-reasoning)} \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \phi) \circ \phi && (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow\ & M, j \Vdash (\rho \wedge (\phi \wedge \psi)) \circ (\phi \wedge \psi) && (M, j \Vdash \psi \text{ and (U8°)})
\end{aligned}
$$

Case 2: $M, w \Vdash \rho \wedge \neg(\phi \wedge \psi) \Rightarrow M, w \Vdash (\rho \wedge \neg\phi) \vee (\rho \wedge \neg\psi)$.
Case 2.1:

$$
\begin{aligned}
& M, w \Vdash (\rho \wedge \neg\phi) \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi) \circ \phi && (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi) \circ (\phi \wedge \psi) && (M, j \Vdash \psi \text{ and (U6)}) \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi \vee \rho \wedge \neg\psi) \circ (\phi \wedge \psi) && \text{(MONU)} \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg(\phi \wedge \psi)) \circ (\phi \wedge \psi) && \text{(Meta-reasoning)}
\end{aligned}
$$

Case 2.2: $M, w \Vdash \rho \wedge \neg\psi$.
Case 2.2.1:

$$
\begin{aligned}
& M, w \Vdash \phi \\
\Rightarrow\ & M^*, w \Vdash \phi && \text{(Filtration Lemma)} \\
\Rightarrow\ & \{w\} = cw^*(w, [\![\phi]\!]_{M^*}) && \text{(CS2)} \\
\Rightarrow\ & w = j && (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow\ & M, j \Vdash \neg\psi && (M, w \Vdash \neg\psi) \\
\Rightarrow\ & \mathit{False} && (M, j \Vdash \psi)
\end{aligned}
$$

Case 2.2.2:

$$
\begin{aligned}
& M, w \Vdash \neg\phi \\
\Rightarrow\ & M, w \Vdash (\rho \wedge \neg\phi) && \text{(Meta-reasoning)} \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi) \circ \phi && (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi) \circ (\phi \wedge \psi) && \text{(U6)} \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg\phi \vee \rho \wedge \neg\psi) \circ (\phi \wedge \psi) && \text{(MONU)} \\
\Rightarrow\ & M, j \Vdash (\rho \wedge \neg(\phi \wedge \psi)) \circ (\phi \wedge \psi) && \text{(Meta-reasoning)}
\end{aligned}
$$

Case 3:

$$
\begin{aligned}
& M, w \Vdash \neg\rho \wedge (\phi \wedge \psi) \\
\Rightarrow\ & M, w \Vdash \neg\rho \wedge \phi && \text{(Meta-reasoning)} \\
\Rightarrow\ & M, j \Vdash (\neg\rho \wedge \phi) \circ \phi && (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow\ & M, j \Vdash (\neg\rho \wedge (\phi \wedge \psi)) \circ \phi && (M, j \Vdash \psi \text{ and (U8°)}) \\
\Rightarrow\ & M, j \Vdash (\neg\rho \wedge (\phi \wedge \psi)) \circ (\phi \wedge \psi) && \text{(U4)}
\end{aligned}
$$

Case 4: $M, w \Vdash (\neg\rho \wedge \neg(\phi \wedge \psi)) \Rightarrow M, w \Vdash (\neg\rho \wedge \neg\phi) \vee (\rho \wedge \neg\psi)$.
Case 4.1:

$$
\begin{array}{ll}
& M, w \Vdash (\neg\rho \wedge \neg\phi) \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi) \circ \phi \qquad (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi) \circ (\phi \wedge \psi) \qquad (M, j \Vdash \psi \text{ and (U6)}) \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi \vee \rho \wedge \neg\psi) \circ (\phi \wedge \psi) \qquad \text{(MONU)} \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg(\phi \wedge \psi)) \circ (\phi \wedge \psi) \qquad \text{(Meta-reasoning)}
\end{array}
$$

Case 4.2: $M, w \Vdash \neg\rho \wedge \neg\psi$.
Case 4.2.1:

$$
\begin{array}{ll}
& M, w \Vdash \phi \\
\Rightarrow & M^*, w \Vdash \phi \qquad \text{(Filtration Lemma)} \\
\Rightarrow & \{w\} = cw^*(w, [\![\phi]\!]_{M^*}) \qquad \text{(CS2)} \\
\Rightarrow & w = j \qquad (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow & M, j \Vdash \neg\psi \qquad (M, w \Vdash \neg\psi) \\
\Rightarrow & \textit{False} \qquad (M, j \Vdash \psi)
\end{array}
$$

Case 4.2.2:

$$
\begin{array}{ll}
& M, w \Vdash \neg\phi \\
\Rightarrow & M, w \Vdash (\neg\rho \wedge \neg\phi) \qquad \text{(Meta-reasoning)} \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi) \circ \phi \qquad (j \in cw^*(w, [\![\phi]\!]_{M^*})) \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi) \circ (\phi \wedge \psi) \qquad \text{(U6)} \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg\phi \vee \rho \wedge \neg\psi) \circ (\phi \wedge \psi) \qquad \text{(MONU)} \\
\Rightarrow & M, j \Vdash (\neg\rho \wedge \neg(\phi \wedge \psi)) \circ (\phi \wedge \psi) \qquad \text{(Meta-reasoning)}
\end{array}
$$

Therefore, by the results of Cases 1–4 and the definition of cw^*, we have that $j \in cw^*(w, [\![\phi \wedge \psi]\!]_{M^*})$. So (CSC) is satisfied.

(NORM) $(\emptyset \not\succ^* X)$. We must show that $\emptyset \succ^* X \Rightarrow$ **False**.

$$
\begin{array}{ll}
& \emptyset \succ^* X \\
\Rightarrow & (\exists \phi, \psi \in \Phi_\chi^+) \\
& \quad (M, w \Vdash \phi \boldsymbol{P} \psi \text{ and} \\
& \quad\; cw^*(w, [\![\phi \wedge \neg\psi]\!]_{M^*}) = \emptyset \text{ and} \\
& \quad\; cw^*(w, [\![\psi \wedge \neg\phi]\!]_{M^*}) = X \text{ and} \\
& \quad\; cw^*(w, [\![\phi \wedge \neg\psi]\!]_{M^*}) \succ^* \\
& \quad\; cw^*(w, [\![\psi \wedge \neg\phi]\!]_{M^*})) \qquad \text{(Definition of } \succ) \\
\Rightarrow & cw^*(w, [\![\bot]\!]_{M^*}) \succ^* \\
& cw^*(w, [\![\psi \wedge \neg\phi]\!]_{M^*}) \qquad (cw^*(w, [\![\bot]\!]_{M^*}) = \emptyset \text{ by (CS1)})
\end{array}
$$

$\Rightarrow cw^*(w, [\![\bot]\!]_{M^*}) \succ^*$
$cw^*(w, [\![\psi \wedge \neg\phi]\!]_{M^*})$ and
$\bot \in \Phi_X^+$ and
$\exists \rho \in \Phi_X^+([\![\rho]\!]_{M^*} = [\![\psi \wedge \neg\phi]\!]_{M^*})$ (Definition of Φ_X^+)
$\Rightarrow cw^*(w, [\![\bot \wedge \neg\rho]\!]_{M^*}) \succ^*$
$cw^*(w, [\![\rho \wedge \neg\bot]\!]_{M^*})$ (Meta-reasoning)
$\Rightarrow M, w \Vdash \bot \boldsymbol{P} \rho$ (Definition of cw^*)
$\Rightarrow$ **False** (N)

The proof for the second part of (NORM), $(X \not\succ^* \emptyset)$, is similar.

(TRAN) $cw^*(w, X \cap \overline{Y}) \succ^* cw^*(w, Y \cap \overline{X})$ and $cw^*(w, Y \cap \overline{Z}) \succ^* cw^*(w, Z \cap \overline{Y}) \Rightarrow cw^*(w, X \cap \overline{Z}) \succ^* cw^*(w, Z \cap \overline{X})$.

$cw^*(w, X \cap \overline{Y}) \succ^* cw^*(w, Y \cap \overline{X})$ and
$cw^*(w, Y \cap \overline{Z}) \succ^* cw^*(w, Z \cap \overline{Y})$
$\Rightarrow \exists\phi\exists\psi\exists\rho(X = [\![\phi]\!]_{M^*}$ and $Y = [\![\psi]\!]_{M^*}$ and
$Z = [\![\rho]\!]_{M^*}$ and $M, w \Vdash (\phi \boldsymbol{P} \psi)$ and
$M, w \Vdash (\psi \boldsymbol{P} \rho)$ and $(\phi, \psi, \rho \in \Phi_X^+))$ (Definition of cw^*)
$\Rightarrow M, w \Vdash \phi \boldsymbol{P} \rho$ and $(\phi, \rho \in \Phi_X^+)$ (TR)
$\Rightarrow cw^*(w, [\![\phi \wedge \neg\rho]\!]_{M^*}) \succ^* cw^*(w, [\![\rho \wedge \neg\phi]\!]_{M^*})$ (Definition of $\succ^*$)
$\Rightarrow cw^*(w, X \cap \overline{Z}) \succ^* cw^*(w, Z \cap \overline{X})$ (Definitions of X, Y, Z)

As a consequence, M^* is an ALX model. Because of the soundness of ALX logic, we know that for any ρ, $\vdash_{\mathrm{ALX}} \rho \Rightarrow \forall w(M^*, w \Vdash \rho)$. This means that ALX also satisfies condition (2) of the finite model property (Definition 9). So ALX has the finite model property. □

Acknowledgement

The authors gratefully acknowledge very helpful suggestions from Maarten Marx, Peter van Emde Boas, Scip Garling, Jaap Kamps, Jean-Jules Meyer and Yao-Hua Tan. This research was supported by a PIONIER grant from the Dutch National Science Foundation (# PGS 50-334).

References

[1] H. Blumer, *Symbolic Interactionism: Perspective and Methods* (Englewood Cliffs, Prentice-Hall, NJ, 1969).

[2] A.L. Brown, S. Mantha and T. Wakayama, Preferences as normative knowledge: towards declarative obligations, in: J.-J.C. Meyer and R.J. Wieringa, eds., *Proceedings DEON'91*, Amsterdam (1991) 142–163.

[3] K. Carley, Efficiency in a garbage can: implications for crisis management, in: J.G. March and R. Weissinger-Baylon, eds., *Ambiguity and Command* (Pitman, Marshfield, MA, 1986) 165–194.
[4] R. Chisholm and E. Sosa, On the logic of "intrinsically better", *Amer. Philos. Q.* **3** (1966) 244–249.
[5] P.R. Cohen and H.J. Levesque, Persistence, intention and commitment, in: M.P. Georgeff and A.L. Lansky, eds., *Proceedings 1986 Workshop on Reasoning about Actions and Plans* (Morgan Kaufmann, San Mateo, CA, 1987) 297–340.
[6] P.R. Cohen and H.J. Levesque, Intention is choice with commitment, *Artif. Intell.* **42** (1990) 213–261.
[7] S. Danielsson, *Preference and Obligation* (Filosofischa föreningen, Uppsala, 1968).
[8] T. Dean and M.P. Wellman, On the value of goals, in: *Proceedings Rochester Planning Workshop*, Rochester, NY (1989).
[9] P.F. Drucker, The theory of business, *Harvard Business Rev.* (Sept.-Oct. 1994) 95–107.
[10] S. French, *Decision Theory, an Introduction to the Mathematics of Rationality* (Ellis Horwood, Chichester, England, 1988).
[11] L.T.F. Gamut, *Logic, Language, and Meaning* (The University of Chicago Press, Chicago, IL, 1990).
[12] A. Giddens, *Central Problems in Social Theory: Action, Structures, and Contradiction in Social Analysis* (University of California Press, Berkeley, CA, 1979).
[13] M.L. Ginsberg, Counterfactuals, *Artif. Intell.* **30** (1986) 35–79.
[14] M.L. Ginsberg and D.E. Smith, Reasoning about action I: a possible worlds approach, *Artif. Intell.* **35** (1988) 165–195; also in: M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann, Los Altos, CA, 1987) 433–463.
[15] G. Grahne, Updates and counterfactuals, in: J. Allen, R. Fikes and E. Sandewall, eds., *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning* (Morgan Kaufmann, San Mateo, CA, 1991) 269–276.
[16] J. Habermas, *The Theory of Communicative Action* (Beacon Press, Boston, MA, 1984).
[17] S. Halldén, *On the Logic of Better*, Library of Theoria **2** (Lund, 1957).
[18] S. Halldén, Preference logic and theory choice, *Synthese* **16** (1966) 307–320.
[19] S. Halldén, *The Foundations of Decision Logic* (CWK Gleerup, Lund, 1980).
[20] J.Y. Halpern and Y. Moses, A guide to completeness and complexity for modal logics of knowledge and belief, *Artif. Intell.* **54** (1992) 319–379.
[21] B. Hansson, Fundamental axioms for preference relations, *Synthese* **18** (1968) 423–442.
[22] S.O. Hansson, A new semantical approach to the logic of preference, *Erkenntnis* **31** (1989) 1–42.
[23] S.O. Hansson, Similarity semantics and minimal changes of belief, *Erkenntnis* **37** (1992) 401–429.
[24] D. Harel, Dynamic logic, in: D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic*, Vol. II (Reidel, Dordrecht, Netherlands, 1984) 497–604.
[25] K. Hirofumi and A. Mendelzon, On the difference between updating a knowledge base and revising it, in: J. Allen, R. Fikes and E. Sandewall, eds., *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning* (Morgan Kaufmann, San Mateo, CA, 1991) 387–394.
[26] Z. Huang, Logics for agents with bounded rationality, ILLC Dissertation series 1994-10, University of Amsterdam (1994).
[27] Z. Huang and M. Masuch, Reasoning about action: a comparative survey, CCSOM Research Report 91-37 (1991).
[28] Z. Huang, M. Masuch and L. Pólos, A preference logic for rational actions, in: R. Blanning and D. King, eds., *Artificial Intelligence in Organization Design, Modeling and Control*, Information Systems Series (IEEE Computer Society Press, forthcoming).
[29] Z. Huang, M. Masuch and L. Pólos, ALX, the x'th action logic for agents with bounded rationality, CCSOM Research Report 92-70a (1992).
[30] Z. Huang, M. Masuch and L. Pólos, ALX2: the quantifier ALX logic, CCSOM Research Report 93-99 (1993).
[31] G.E. Hughes and M.J. Cresswell, *A Companion to Modal Logic* (Methuen, New York, 1984).
[32] F. Jackson, ed., *Conditionals* (Oxford University Press, 1991).
[33] P. Jackson, On the semantics of counterfactuals, in: *Proceedings IJCAI-89*, Detroit, MI (1989).
[34] R.C. Jeffrey, *The Logic of Decision* (New York, 2nd ed., 1983).
[35] S. Kambhampati and S. Kedar, A unified framework for explanation-based generalization of partially ordered and partially instantiated plans, *Artif. Intell.* **67** (1994) 29–70.

[36] H. Katsuno and A.O. Mendelzon, Propositional knowledge base revision and minimal change, *Artif. Intell.* **52** (1991) 263–294.
[37] A. Kron and V. Milovanović, Preference and choice, *Theory Decision* **6** (1975) 185–196.
[38] D.K. Lewis, *Counterfactuals* (Blackwell, Oxford, 1973).
[39] N. Luhmann, *The Differentiation of Society* (Columbia University Press, New York, 1982).
[40] J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of AI, in: B. Meltzer and D. Michie, eds., *Machine Intelligence* **4** (Edinburgh University Press, 1969).
[41] J.G. March, The technology of foolishness, in: J.G. March and J.P. Olsen, eds., *Ambiguity and Choice in Organizations* (Bergen, Norway, Universitetsforlaget, 1976) 69–81.
[42] J.G. March and J.P. Olsen, Garbage can models of decision making in organizations, in: J.G. March and R. Weissinger-Baylon, eds., *Ambiguity and Command* (Pitman, Marshfield, MA, 1986) 11–53.
[43] N. Martí-Oliet and J. Meseguer, Action and change in rewriting logic, SRI-CSL Tech. Report 94-07 (1994).
[44] M. Masuch, Formalization of Thompson's organization in action, CCSOM Research Report 91-32 (1991).
[45] M. Masuch, Introduction, in: M. Masuch and M. Warglien, eds., *Artificial Intelligence in Organization and Management, Models of Distributed Activity* (Elsevier-North Holland, Amsterdam, 1992) 1–19.
[46] M. Masuch and Z. Huang, A logical deconstruction of organizational action: formalizing Thompson's *Organizations in Action* into a multi-agent action logic, CCSOM Working Paper 94-120.
[47] J.-J.C. Meyer, Using programming concepts in deontic reasoning, in: R. Bartsch, J. van Benthem and P. van Emde Boas, eds., *Semantics and Contextual Expression* (Foris Publications, Dordrecht, Netherlands, 1989) 117–145.
[48] N.J. Moutafakis, *The Logic of Preference* (Reidel, Dordrecht, Netherlands, 1987).
[49] J.D. Mullen, Does the logic of preference rest on a mistake, *Metaphilosophy* **10** (1979) 247–255.
[50] D. Nute, Conditional logic, in: D. Gabby and F. Guenthner, eds., *Handbook of Philosophical Logic*, Vol. II (1986) 387–439.
[51] J.F. Padgett, Managing garbage can hierarchies, *Administrative Sci. Q.* **25** (1980) 583–604.
[52] T. Parsons, *The Structure of Social Action* (Free Press, Glencoe, IL, 1937).
[53] T. Parsons, *The Social System* (Routledge and Kegan Paul, London, 1951).
[54] M.E. Pollack, The uses of plans, *Artif. Intell.* **57** (1992) 43–68.
[55] J.L. Pollock, *Subjunctive Reasoning* (Reidel, Dordrecht, Netherlands, 1976).
[56] J.L. Pollock, A refined theory of counterfactuals, *J. Philos. Logic* **10** (1981) 239–266.
[57] L. Pólos, Updated situation semantics, *J. Symbolic Logic* **58** (1993) 1143–1144.
[58] L. Pólos and M. Masuch, Information states in situation semantics, in: L. Pólos and M. Masuch, eds., *Applied Logic: How, What and Why* (Kluwer Academic Publishers, Dordrecht, Netherlands, to appear) 177–218.
[59] A.S. Rao and M.P. Georgeff, Modeling rational agents within a BDI-architecture, in: J. Allen, R. Fikes and E. Sandewall, eds., *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning* (Morgan Kaufmann, San Mateo, CA, 1991) 473–484.
[60] N. Rescher, Semantic foundations for the logic of preference, in: N. Rescher, ed., *The Logic of Decision and Action* (University of Pittsburgh Press, Philadelphia, PA, 1967).
[61] A. Schutz, *The Phenomenology of the Social World* (Northwestern University Press, Evanston, IL, 1967).
[62] D. Scott, Towards a mathematical theory of computation, in: *Proceedings Fourth Annual Princeton Conference on Information Science and Systems* (1970) 169–176.
[63] D. Scott, Domains for denotational semantics, in: M. Nielsen and E.T. Schmidt, eds., Lecture Notes in Computer Science **140** (Springer-Verlag, Berlin, 1982) 577–613.
[64] K. Segerberg, The logic of deliberation action, *J. Philos. Logic* **11** (1982) 233–254.
[65] H.A. Simon, A behavioral model of rational choice, *Q. J. Economics* **69** (1955) 99–118.
[66] H.A. Simon, On the concept of organizational goal, *Administrative Sci. Q.* **9** (1964) 1–22.
[67] H.A. Simon, Bounded rationality, in: J. Eatwell et al., eds., *The New Palgrave* (Macmillan, London, 1987).
[68] R. Stalnaker, A theory of conditionals, in: *Studies in Logical Theory, Amer. Philos. Q.* **2** (1968) 98–122.
[69] J.D. Thompson, *Organizations in Action, Social Science Bases of Administrative Theory* (McGraw-Hill, New York, 1967).

[70] M.P. Wellman and J. Doyle, Preferential semantics for goals, in: *Proceedings AAAI-91*, Anaheim, CA (AAAI Press, 1991) 698–703.
[71] M. Winslett, Reasoning about action using a possible models approach, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 89–93.
[72] G.H. von Wright, *The Logic of Preference* (Edinburgh, 1963).
[73] G.H. von Wright, The logic of preference reconsidered, *Theory Decision* **3** (1972) 140–169.