The Logic of Permission and Obligation in the Framework of ALX.3: how to avoid the paradoxes of deontic logics.

Zhisheng Huang and Michael Masuch
Center for Computer Science in Organization and Management (CCSOM)
University of Amsterdam
Sarphitistraat 143
1018 GD Amsterdam
The Netherlands
email: {huang,michael}@ccsom.uva.nl

Abstract

Standard deontic logic features fairly serious socalled "paradoxes" (technically: counterintuitive validities). Much energy in deontic logic has been spendt on avoiding these "paradoxes". We suggest a reformulation of deontic logic in terms of a multi-agent logic, ALX.3, where a "super-agent" (think of a legislature) lays down the law of the land and other agents have to follow its rules. In particular, obligations are reformulated in terms of "preferences" of the superagent. We can show that our approach avoids the classical "paradoxes" of deontic logic, thanks to the properties of the preference operator of ALX.3.

1 Introduction

Deontic logic is a branch of modal logic for reasoning about social norms by means of modal operators denoting states of *obligation* (written as **O**), *permission*(written as **P**), and *prohibition* (written as **F**, from "forbidden"). Deontic logic has many potential application in areas such as law, computer science, and sociology, but the standard versions of deontic logic suffer from serious "paradoxes"—pardoxes not in a technical sense but in the sense of counterintuitive validities—that have kept deontic logic from living up to its full potential. Here are some examples of such validities:

- Ross' Paradox: $\mathbf{O}\phi \rightarrow \mathbf{O}(\phi \vee \psi)$.
  This validity would justify propositions such as: "if one is obliged to mail the letter, one is obliged to either mail the letter or burn it".
- Penitent's Paradox: $\mathbf{F}\phi \rightarrow \mathbf{F}(\phi \wedge \psi)$.
  This validity would justify propositions such as: "it it is forbidden to commit a crime, then it is also forbidden to commit a crime and do penitence for it".
- Good Samaritan Paradox: $\phi \rightarrow \psi \Rightarrow \mathbf{O}\phi \rightarrow \mathbf{O}\psi$.
  This validity implies, for example, that the fact that a good Samaritian help a victim if the victim has been robbed, together with the obligation to help vitims after a robbery implies the obligation to rob the victim in the first place.

The underlying reason for these paradoxes is technical: at least one of the modal operators is introduced as a primitive modality in the context of a normal modal logic (roughly speaking: a logic with Kripke-style semantics and no absurd worlds), and such operators come with the standard properties. The Good Samaritian paradox is a consequence of the monotonicity of normal modal logics, while Ross's paradox is just a special case of this monotonicity with the appropriate substitutions made for the validity $(\phi \to (\phi \vee \psi))$ (normal modal logics always have the propositional tautologies as validities). Penitent's paradox, is, in, fact the dual version of Ross's paradox with the obligation operator replaced by the prohibition operator.

In fact, not only deontic logics suffer from the side effects of normality; similar problems arise in epistemic logics, where the epistemic operators are also acting as normal modalities (logical omniscience is a typical side-effect of normality in epistemic logics).

Various attempts have been made to circumvent the paradoxes. Anderson, for example [1], define the the prohibition operator with the help of a propositional constant V as follows:

$$\mathbf{F}\phi \stackrel{\text{def}}{\Longleftrightarrow} \Box(\phi \to \mathbf{V}).$$

where the meaning of the constant V is "liablity to sanction or punishment", and the box-operator assumes the standard meaning of alethic modal logic: a state ϕ is forbidden a if and only if the state ϕ necessarily impies the sanction for the agent. As it turned out however, Anderson's approach missed its goal [10]. The Good Samaritan Paradox does not go away, and neither do Ross' and Penitent's paradoxes—being special cases of the former one.

Inspired by Anderson's reduction to alethic modal logic, J.-J. Meyer proposes another solution in [9], There, Meyer in this case to propositional dynamic logic, still using Anderson's special violation atom V. One of the consequences of the use of dynamic logic is the distinction between propositions and actions. In Meyer's approach, α is forbidden ($\mathbf{F}$), permitted ($\mathbf{P}$), and obligated ($\mathbf{O}$) are reduced to dynamic expressions as: $\mathbf{F}\alpha \stackrel{\text{def}}{\Longleftrightarrow} [\mathbf{a}]\mathbf{V}$, $\mathbf{P}\alpha \stackrel{\text{def}}{\Longleftrightarrow} \neg\mathbf{F}\alpha$, and $\mathbf{O}\alpha \stackrel{\text{def}}{\Longleftrightarrow} \mathbf{F}(\neg\alpha)$. As it is pointed out in [10], Ross' paradox remains in Meyer's reformulation.

Similarly, inspired by Anderson's idea of reformulation of deontic logics, in this paper, we propose an approach of reformulation of deontic logics by using preference operator instead of the logical implication or dynamic actions. In other words, we try to define the forbidden operator in terms of preference operator in a multiple agent logic. We assume that there exists a superagent in the multiple agent environment. Informally, a state is forbidden for an agent i if and only if the agent's favourite on the state is against the superagent's. We find that the preference operator in ALX logics is suitable to fulfill such a task, where ALX logics are ones proposed by the authors in [5, 7]. So far there are three version of ALX logics: ALX.1 , ALX.2, and ALX.3. ALX.1 is a propositional action logic for agents with bounded rationality. ALX.2 is an ALX with first order language description. ALX.3 is a multi-agent action logic wth the first order language description.

In the following, this paper is organized as follows: First, in section 2, we briefly overview the preference operator in ALX and its semantics, discuss the formal properties of preference, and define the goodness and badness operators in terms of the preference operators. Then, in the section 3, we reformulate the deontic logic in terms of the goodness and badness operators, discuss the formal properties of this new deontic logic, show how those typical "paradoxes" can be avoided in this logic, and make comparison with other approaches to discuss our solution and its advantage. Finally, in section 4, we make concluding remarks and discuss the future directions.

2 ALX logics and its Preference Operator

2.1 ALX Logics

In [5, 7], 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 *mnimal change*, and more recent approaches to *update semantics*. ALX (the x's 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. ALX.3 is a multiple agent version of ALX logics.

In this paper, we use only the preference operator to reformulate the deontic logic. In the following, we focus on the preference operator and its semantics. In particular, we focus on the preference operator in ALX.3. The ALX.3 logic is discussed in details in [8]. A short overview is given in the appendix. In the following, we use a simplized version of ALX.3, namely, a multiple agent preference logic. It is easy to see how we can reduce ALX.3 into this simplized logic we need in this paper.

2.2 Preference

Preferences provide the basis for rational action in ALX. Following von Wright [12], 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 thetates 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, which is called *the conjunction expansion principle.* Preferences are expressed via two-place modal operators; if the agent prefers the proposition ϕ to the proposition ψ, we write $\phi \boldsymbol{P}_i \psi$.

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 [11]. 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.

Let W be the set of all possible worlds in a semantic model M. Semantically, a closest world function *cw* is a function $W \times \mathcal{P}(\mathcal{W}) \rightarrow \mathcal{P}(\mathcal{W})$, which assigns a set of possible worlds to each world possible with respect to a set of possible world. In other words,

$cw(w, [\![\phi]\!]^v_M) = [\![\psi]\!]^v_M$ means that $[\![\psi]\!]^v_M$ is the set of the closest-ϕ-world to the world w, where $[\![\phi]\!]^v_M$ as usual is a set of world in which ϕ holds, i.e. $[\![\phi]\!]^v_M = \{w \in W : M, w, v \models \phi\}$.

The semantic component of the preference in the model is a function $\succ: AGENT \to \mathcal{P}(\mathcal{P}(W) \times \mathcal{P}(W))$, which assigns a comparison relation for preferences to each agent.

Moreover, in ALX.3 semantic models, $\succ$ satisfy the following conditions

$$\begin{array}{ll} \boldsymbol{(NORM)}: & (\emptyset \not\succ_i X), (X \not\succ_i \emptyset), \text{where } \succ_i = \succ(i) \text{ for each agent } i \in AGENT. \\ \boldsymbol{(TRAN)}: & cw(w, X \cap \overline{Y}) \succ_i cw(w, Y \cap \overline{X}) \text{ and } cw(w, y \cap \overline{Z}) \succ_i cw(w, Z \cap \overline{Y}) \\ & \Rightarrow cw(w, X \cap \overline{Z}) \succ_i cw(w, Z \cap \overline{X}), \text{where } \overline{X} = W - X. \end{array}$$

(NORM) and (TRAN) constrain the semantic preference relation. (NORM) is required in support of the logical axiom (N) (normality), which, in turn, protects the preference logic against counterintuitive consequences. (TRAN) guarantees the soundness of the logical axiom (TR) which, in turn, assures transitivity for preferences.

The meaning function of the preference is:

$$M, w, v \models \phi \boldsymbol{P}_i \psi \text{ iff } cw(w, [\![\phi \wedge \neg\psi]\!]^v_M) \succ_i cw(w, [\![\psi \wedge \neg\phi]\!]^v_M).$$

The interpretation of $\phi \boldsymbol{P}_i \psi$ assures the conjunction expansion principle.

2.3 Formal Properties of Preference

The preference operator in this semantics has the following axioms and inference rules:

Axioms

(CEP) $\phi \boldsymbol{P}_i \psi \leftrightarrow (\phi \wedge \neg\psi)\boldsymbol{P}_i(\neg\phi \wedge \psi)$
(N) $\neg(\bot \boldsymbol{P}_i \phi), \neg(\phi \boldsymbol{P}_i \bot)$
(TR) $(\phi \boldsymbol{P}_i \psi) \wedge (\psi \boldsymbol{P}_i \chi) \to (\phi \boldsymbol{P}_i \chi)$

Inference Rules

$(MP) \vdash \phi \;\&\; \vdash \phi \to \psi \Rightarrow \vdash \psi$
$(SUBP) \vdash (\phi \leftrightarrow \phi') \;\&\; \vdash (\psi \leftrightarrow \psi') \Rightarrow \vdash (\phi \boldsymbol{P}_i \psi) \leftrightarrow (\phi' \boldsymbol{P}_i \psi')$

(CEP) states the conjunction expansion principle. (N) establishes "normality" and (TR) transitivity. As noted before, (TR) would go if its semantic equivalent, (TRAN), goes, so we could have non-transitive preferences. (CEP) and (N) together imply the irreflexivity and contraposition (CP) of the $\boldsymbol{P}$ operator[7]. We have the modus ponens (MP) for obvious reasons. Furthermore, logically equivalent propositions are substitutable in preference formulae (SUBP). Note that we do *not* have monotonicity for preferences. Because of this, we are able to avoid the counterintuitive deductive closure of goals that mars other action logics.

Furthermore, preferences in this semantics have pleasant logical properties. In particular, the preference operator can *avoid* the following counterintuitive properties.

- Necessitation rule for preference:
  $\models \phi \Rightarrow \models \phi \boldsymbol{P}_i \psi$ and $\models \phi \Rightarrow \models \psi \boldsymbol{P}_i \phi$
  The first half of this property exmplified by the statement " if it is necessary that the sun rises in mornings, then the state of the sun's rising in the morning is always preferred to any state else". It is definitely counterintuitive. It is also easy to find the counter-example for the second half of this property.

- Closure for preference:
  $\models (\phi \to \psi) \Rightarrow \models (\phi \boldsymbol{P}_i \phi' \to \psi \boldsymbol{P}_i \phi')$ and $\models (\phi \to \psi) \Rightarrow \models (\phi' \boldsymbol{P}_i \phi \to \phi' \boldsymbol{P}_i \psi)$
  This property means that if I prefer (having) tea to (having) coffee, then I prefer (having) tea or one million dollars to (having) coffee, since having tea always implies having tea or one million dollars.

- Conjunction extension:
  $\models \phi \boldsymbol{P}_i \psi \to (\phi \wedge \phi') \boldsymbol{P}_i \psi$ a nd $\models \psi \boldsymbol{P}_i \phi \to \psi \boldsymbol{P}_i (\phi \wedge \phi')$
  This property has a consequence that if I prefer tea to coffee, then I prefer tea and poison to coffee.

- Disjunction extension
  $\models \phi \boldsymbol{P}_i \psi \to (\phi \vee \phi') \boldsymbol{P}_i \psi$ and $\models \psi \boldsymbol{P}_i \phi \to \psi \boldsymbol{P}_i (\phi \vee \phi')$
  This property is a simplized case of the closure for preferences. We can use the same counter-example for this property.

2.4 Good and Bad States

Following von Wright, we define a "good" state ϕ as a state that agent i prefers to its negation, and converselyfor a bad state:

$$\mathbf{Good}_i(\phi) \overset{\text{def}}{\iff} (\phi \boldsymbol{P}_i \neg\phi) \qquad \mathbf{Bad}_i(\phi) \overset{\text{def}}{\iff} (\neg\phi \boldsymbol{P}_i \phi).$$

Proposition 1 (More properties of goodness and badness)
(a) $\phi \boldsymbol{P}_i \psi \wedge \mathbf{Good}_i \psi \to \mathbf{Good}_i \phi$.
(b) $\phi \boldsymbol{P}_i \psi \wedge \mathbf{Bad}_i \phi \to \mathbf{Bad}_i \psi$.
(c) $\mathbf{Good}_i \phi \leftrightarrow \mathbf{Bad}_i \neg\phi$.
(d) $\mathbf{Good}_i \phi \to \neg\mathbf{Bad}_i \phi$.
(e) $\mathbf{Bad}_i \phi \to \neg\mathbf{Bad}_i \neg\phi$.

PROOF: (a)

$$\begin{array}{lll} & \vdash \phi \boldsymbol{P}_i \psi \wedge \mathbf{Good}_i \psi & \\ \Rightarrow & \vdash \phi \boldsymbol{P}_i \psi \wedge \psi \boldsymbol{P}_i \neg\psi & \\ \Rightarrow & \vdash \phi \boldsymbol{P}_i \psi \wedge \phi \boldsymbol{P}_i \neg\psi & \text{(TR)} \\ \Rightarrow & \vdash \phi \boldsymbol{P}_i \psi \wedge \psi \boldsymbol{P}_i \neg\phi & \text{(CP)} \\ \Rightarrow & \vdash \phi \boldsymbol{P}_i \neg\phi & \text{(TR)} \\ \Rightarrow & \vdash \mathbf{Good}_i \phi & \end{array}$$

The proof for (b) is similar to the proof of (a). (c)-(e) are straightforward from the definitions. □

The notion of the goodness and badness are crucial notions which are used to define the deontic operators in this paper. Since these two operators are defined in terms of the preference operator, they also can avoid those counter-intuitive properties of preferences.

3 Deontic logic in the framework of ALX.3

3.1 Defining the operations of forbiding, obligation, and permission

ALX.3 is a multi-agent action logic. We can assume that there exists a super-agent, written *sg* among the multi agents. This super-agent need not be a dictator. She could be something like legislature. She lays down the law of land according to her preferences. Therefore, we can define the **F** operator ("forbiding") as follows:

$$\mathbf{F_i}\phi \overset{\text{def}}{\Longleftrightarrow} \mathbf{Bad_{sg}Good_i}\phi.$$

(A state is forbidden for an agent i if and only if the agent's considering about the state ϕ is against the super-agent.)

Furthermore, we define the obligation operator and permission operator in the standard way, namely,

$$\mathbf{O_i}\phi \overset{\text{def}}{\Longleftrightarrow} \mathbf{F_i}(\neg\phi),$$

$$\mathbf{P_i}\phi \overset{\text{def}}{\Longleftrightarrow} \neg\mathbf{O_i}(\neg\phi).$$

Adding the above two definitions into ALX.3, we obtain the deontic ALX, called DALX. Since ALX.3 is sound and complete [7], adding more definitions actually would not change any formal properties of the original logic. Therefore, DALX is sound and complete as well. Furthermore, DALX keeps a lot of nice validities in ordinary deontic logics.

The following are the theorems of the deontic ALX:

(a) Consistency of "forbidden"
$\mathbf{F_i}\phi \to \neg\mathbf{F_i}\neg\phi$.

(b) Consistency of obligation
$\mathbf{O_i}\phi \to \neg\mathbf{O_i}\neg\phi$.

(c) Connection between forbidden and permitted states
$\mathbf{P_i}\phi \leftrightarrow \neg\mathbf{F_i}\phi$.

(d) Obligation implies permission
$\mathbf{O_i}\phi \to \mathbf{P_i}\phi$.
PROOF: $\mathbf{O_i}\phi \Rightarrow \mathbf{F_i}\neg\phi \Rightarrow \neg\mathbf{F_i}\phi \Rightarrow \mathbf{P_i}\phi$. □

(e) "Forbidden" is not permitted
$\mathbf{F_i}\phi \leftrightarrow \neg\mathbf{P_i}\phi$.

(f) No contradictory obligation
$\neg\mathbf{O_i}(\phi \wedge \neg\phi)$.

(g) Permission property
$\neg\mathbf{P_i}\phi \to \mathbf{P_i}\neg\phi$
PROOF: $\neg\mathbf{P_i}\phi \Rightarrow \mathbf{F_i}\phi \Rightarrow \neg\mathbf{F_i}\neg\phi \Rightarrow \mathbf{P_i}\neg\phi$. □

(h) Substitution rule for "forbidden"
$\models (\phi \leftrightarrow \psi) \Rightarrow \models (\mathbf{F_i}\phi \leftrightarrow \mathbf{F_i}\psi)$.

(i) Substitution rule for obligation
$\models (\phi \leftrightarrow \psi) \Rightarrow \models (\mathbf{O_i}\phi \leftrightarrow \mathbf{O_i}\psi)$.

(j) Substitution rule for permission
$\models (\phi \leftrightarrow \psi) \Rightarrow \models (\mathbf{P_i}\phi \leftrightarrow \mathbf{P_i}\psi)$.

3.2 Avoidance of Paradoxes

This new deontic logic not only preserve a lot of nice validities of ordinary deontic logics. More importantly, it can avoid the paradoxes. Just observe the connection between those paradoxes and the counterintuitive properties:

- Ross's paradox is an example of the disjunction extension property.
- Penitent's paradox is an example of the conjunction extension property.
- The Good Samaritan paradox is an example of the closure under logical implication.

Since all three deontic operations in DALX are defined in terms of the preferences, those paradoxes are avoided.

Proposition 2 *The deontic logic ALX logic DALX can avoid (i) Ross's paradox, (ii) Penitent's paradox, and (iii) Good Samaritan paradox.*

PROOF: For more technical details of the proof, just see the appendix 2. □

A natural question at this moment is that how about other paradoxes, since the above three paradoxes are not exclusive. In the following, we would like to examine some of them. Actually, you can see that most of them can be reduced the above three typical paradoxes.

- Derived Obligation 1: $\mathbf{O}\phi \rightarrow \mathbf{O}(\psi \rightarrow \phi)$.
  The derived obligation 1 is logically equal to the expression $\mathbf{O}\phi \rightarrow \mathbf{O}(\neg\psi \vee \phi)$. Since the formula ψ in this expression is arbitrary, the derived obligation actually is Ross' Paradox. Therefore, our approach can avoid this paradox as well.
- Derived Obligation 2: $\mathbf{O}\neg\phi \rightarrow \mathbf{O}(\phi \rightarrow \psi)$
  Similarly, this paradox can be reduced into the expression $\mathbf{O}\neg\phi \rightarrow \mathbf{O}(\neg\phi \vee \psi)$. Again, it actually is Ross' Paradox.
- Chisholm's Paradox: $\mathbf{O}\phi \wedge \mathbf{O}(\phi \rightarrow \psi) \wedge (\neg\phi \rightarrow \mathbf{O}\neg\psi) \wedge \neg\phi) \equiv \mathbf{false}$.
  Here are Chisholm's contrary -to-duty imperatives: i) it ought to be that someone goes to the assistance of his neighbours, ii) it ought to be that if he does go he

tell them he is coming, iii) if he does not go then he ought not to tell them he is coming, and iv) he does not go. However, in ordinary deontic logics, the above four statements together implies a contradiction. The reasoning is as follows: From $(\neg\phi \to O\neg\psi) \wedge \neg\phi$, we have $\mathbf{O}\neg\psi$, which is valid in any logic. Then, by K-axiom $\mathbf{O}\phi \wedge \mathbf{O}(\phi \to \psi)$, we have $\mathbf{O}\psi$. So the contradiction comes. The K-axiom is valid in any standard Kripke semantics. Fortunately, there is no K-axiom for the preference in ALX.3. Therefore, $DALX$ can avoid Chisholm's Paradox.

3.3 Comparison

Anderson's reformulation cannot avoid most paradoxes, since this approach uses the logical implication to define the deontic operators. The monotonicity rule is valid in his logic. However, the monotonicity rule is exactly the Good Samaritan Paradox. Ross' Paradox and Penitent Paradox are just special cases of the Good Samaritan Paradox. Furthermore, Derived Obligations are just special cases of Ross Paradox. Therefore, Anderson's approach cannot avoid Derived Obligations. Since K-axiom is valid in Anderson's deontic logic, this approach cannot avoid Chisholm's Paradox as well.

Meyer's approach uses dynamic actions instead of the logical implication to define the deontic operators. This approach can get rid of most of the nasty paradoxes including Chisholm's Paradox. Furthermore, some of these paradoxes are not even expressible in the language any more, such as the Derived Obligation $\mathbf{O}\neg\phi \to \mathbf{O}(\phi \to \psi)$. However, Ross Paradox remains in Meyer's approach, since this approach use the dynamic actions, and in any standard dynamic logic, the axiom $[\alpha]\phi \to [\alpha \cup \beta]\phi$ is always valid. Furthermore, although $\mathbf{F}\alpha \to \mathbf{F}(\alpha \wedge \beta)$ is not expressible any more, however, $\mathbf{F}\alpha \to \mathbf{F}(\alpha \& \beta)$ is still one of the theorems of this logic, where & means a parallel composition. Therefore, Penitent Paradox remains, too.

The proposed reformulation in this paper use neither the logical implication nor the dynamic actions but preferences. Furthermore, the preference operator in ALX has a lot of pleasant properties. It has no the monotonicity rule, no K-axiom, etc. These are the main advantages why this approach can avoid those nasty paradoxes.

4 Concluding Remarks

We have propose a deontic logic, DALX, in terms of ALX logic without introducing any new primitive operator. The main idea of this new deontic logic is that we assume there exists a super-agent among the multi-agents, and define the deontic operators for an ordinary agent i in terms of the superagent's and the agent's preferences. Thanks to the properties of the preference operator in ALX, DALX avoids the paradoxes of ordinary deontic logics.

Although the approach depends on the super-agent existence assumption, we do not intend to commit any theological ontology, say, the super-agent is God. Moreover, the super-agent need not be a dictator. As a matter of fact, the assumption is purely conventional. The super-agent may have different interpretations, the super-agent may be legislature, or anything else which has the highest priority to establish the norms for agents.

In this paper, we have not yet offered any theory of the super-agent. That is one of interesting further work. Once we have any theory of the super-agent, we can capture more insights of the deontic operators under the framework of DALX.

5 Appendix 1: Formal Syntax and Semantics of ALX.3

5.1 Formal Syntax

ALX.3 has a sorted first-order description language. There are predicate letters, regular variables, variables reserved for agent and actions respectively, plus the corresponding constant letters:

(1) For each natural number n (≥ 1), a countable set of n-place predicate letters, PRE_n, written as $p_i, p_j, ...$

(2.1) A countable set of regular variables, $RVAR$, written as $x, x_1, y, z, ...$

(2.2) A countable set of action variables, $AVAR$, written as $a, a_1, b, ...$

(2.3) A countable set of agent variables, $AGVAR$, written as $i, i_1, j, ...$

(3.1) A countable set of regular constants, $RCON$, written as $c, c_1, c_2, ...$

(3.2) A countable set of actions constants, $ACON$, written as $ac, ac_1, ac_2, ...$

(3.3) A countable set of agent constants, $AGCON$, written as $ag, ag_1, ag_2, ...$

Furthermore, we have the usual booleans, an existential quantifier, a unary operator for beliefs, binary operators for preferences and causality, a dynamic operator type for actions, and operators that establish a sequence of, or indeterminate choice between, actions. Finally, there are comma and brackets:

(4) The symbols $\neg$(negation), $\wedge$(conjunction), $\boldsymbol{B}$(belief), $\exists$ (existential quantifier), $\boldsymbol{P}$(preference), $\leadsto$(conditional), ;(sequence), $\cup$(choice), $\langle, \rangle$, (, and).

Definition 1 (Variable) *Define the set of variables VAR as follows:*
$VAR = RVAR \cup AVAR \cup AGVAR$.

Definition 2 (Constant) *define the set of constants CON as follows:*
$CON = RCON \cup ACON \cup AGCON$.

Definition 3 (Term) *Define the set of terms $TERM$ as follows:*
$TERM = VAR \cup CON$.

Definition 4 (Action Term) *Define the set of action terms $ATERM$ as follows:*
$ATERM = AVAR \cup ACON$.

Definition 5 (Agent Term) *Define the set of agent terms $AGTERM$ as follows:*
$AGTERM = AGVAR \cup AGCON$.

We use $t, t_1, ...,$ to denote terms, $a, a_1, ...,$ to denote action terms, $i, j, ...,$ to denote agent terms, if that does not cause any ambiguity.

An atomic first-order formula is defined as usual:

Definition 6 (ATOM) *Define the set of atomic formulas $ATOM$ as follows:*
$ATOM =_{df} \{p(t_1, t_2, ..., t_n) \ : \ p \in PRE_n, t_1, t_2, ..., t_n \in TERM\}$

Primitive actions carry an agent index. Compound actions need not carry an agent index; this allows for the sequencing of actions carried out by different agents:

Definition 7 (ACTION) *Define the set of action expressions $ACTION$ recursively as follows:*

- $a \in ATERM, i \in AGTERM \Rightarrow a_i \in ACTION.$
- $a, b \in ACTION \Rightarrow (a; b), (a \cup b) \in ACTION.$

The definition of formulas is standard:

Definition 8 (FORMULA) *Define the set of formulae* **FML** *recursively as follows:*

- $ATOM \subseteq FML.$
- $\phi \in FML \Rightarrow \neg\phi \in FML.$
- $\phi, \psi \in FML \Rightarrow (\phi \wedge \psi) \in FML.$
- $\phi \in FML, x \in VAR \Rightarrow (\exists x \phi) \in FML.$
- $\phi \in FML, a \in ACTION \Rightarrow (\langle a \rangle \phi) \in FML.$
- $\phi, \psi \in FML \Rightarrow (\phi \rightsquigarrow \psi) \in FML.$
- $\phi, \psi \in FML, i \in AGTERM \Rightarrow (\phi \boldsymbol{P}_i \psi) \in FML.$
- $\phi \in FML, i \in AGTERM \Rightarrow \boldsymbol{B}_i \phi \in FML.$

5.2 Semantics

Definition 9 (ALX.3 Model) *Call*

$$M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I \rangle$$

an ALX.3 model, if

- **O** *is a set of objects,*
- PA *is a set of primitive actions,*
- $AGENT$ *is a set of agents,*
- W *is a set of possible worlds,*
- $cw : W \times \mathcal{P}(W) \rightarrow \mathcal{P}(W)$ *is a closest world function,*
- $\succ : AGENT \rightarrow \mathcal{P}(\mathcal{P}(W) \times \mathcal{P}(W))$ *is a function that assigns a comparison relation for preferences to each agent,*

- $\mathcal{R}: AGENT \times PA \rightarrow \mathcal{P}(W \times W)$ *is a function that assigns an accessibility relation to each agent and each primitive action,*

- $\mathcal{B}: AGENT \rightarrow \mathcal{P}(W \times W)$ *is a function that assigns an accessibility relation for the belief operation to each agent,*

- I *is a pair* $\langle I_P, I_C \rangle$,
where I_P *is a predicate interpretation function that assigns to each* n*-place predicate letter* $p \in PRE_n$ *and each world* $w \in W$ *a set of* n *tuples* $\langle u_1, ..., u_n \rangle$*, where each of the* $u_1, .., u_n$ *is in* $D = O \cup PA \cup AGENT$*, called a domain, and*
I_C *is a constant interpretation function that assigns to each regular constants* $c \in RCON$ *an object* $d \in O$*, assigns to each action constant* $ac \in ACON$ *a primitive action* $a_p \in PA$*, and assigns to each agent constant* $g \in AGCON$ *an agent* $a_g \in AGENT$.

and if $cw, \succ, \mathcal{B}$ *satisfy the following conditions respectively:*

$$\begin{array}{ll}
(\boldsymbol{CS1}): & cw(w, X) \subseteq X. \\
(\boldsymbol{CS2}): & w \in X \Rightarrow cw(w, X) = \{w\}. \\
(\boldsymbol{CS3}): & cw(w, X) = \emptyset \Rightarrow cw(w, Y) \cap X = \emptyset. \\
(\boldsymbol{CS4}): & cw(w, X) \subseteq Y \text{ and } cw(w, Y) \subseteq X \Rightarrow cw(w, X) = cw(w, Y). \\
(\boldsymbol{CS5}): & cw(w, X) \cap Y \neq \emptyset \Rightarrow cw(w, X \cap Y) \subseteq cw(w, X).
\end{array}$$

For each agent $i \in AGENT$,

$$\begin{array}{ll}
(\boldsymbol{NORM}): & (\emptyset \not\succ_i X), (X \not\succ_i \emptyset), \text{ where } \succ_i = \succ(i). \\
(\boldsymbol{TRAN}): & cw(w, X \cap \overline{Y}) \succ_i cw(w, Y \cap \overline{X}) \text{ and } cw(w, y \cap \overline{Z}) \succ_i cw(w, Z \cap \overline{Y}) \\
 & \Rightarrow cw(w, X \cap \overline{Z}) \succ_i cw(w, Z \cap \overline{X}), \text{ where } \overline{X} = W - X.
\end{array}$$

$$\begin{array}{ll}
(\boldsymbol{SEB}): & \forall w \exists w' (\langle w, w' \rangle \in \mathcal{B}_i), \text{ where } \mathcal{B}_i = \mathcal{B}(i). \\
(\boldsymbol{TRB}): & \langle w, w' \rangle \in \mathcal{B}_i \text{ and } \langle w', w'' \rangle \in \mathcal{B}_i \Rightarrow \langle w, w'' \rangle \in \mathcal{B}_i
\end{array}$$

(CS#) constrain the closest world function. Note that we do not require uniqueness for closest world. (NORM) and (TRAN) constrain the semantic preference relation. (NORM) is required to support the logical axiom (N) (normality), which, in turn, protects the preference logic against counterintuitive consequences. Its direct effect is to rule out the occurrence of $\perp$ in preference statements. By conjunction expansion principle, (NORM) actually implies (IRE): $\neg(\phi \boldsymbol{P}_i \phi)$, [1] that requires irreflexivity, since we are working with a "strong" preference. (TRAN) guarantees the soundness of the logical axiom (TR) which, in turn, assures transitivity for preferences. (SEB) establishes the seriality of the beliefs and prevents agents from believing $\perp$, while (TRB) assures positive introspection. If (TRB) would go, the logical axiom (4B) would go as well, so we could have a version of ALX without positive introspection. (SEB) and (TRB) make the relation $\mathcal{B}$ serial and transitive. They are standard requirements for the semantics of beliefs.

Definition 10 (Valuation of Variables) *A valuation of variables* v *in the domain* D *of an ALX.3 model* M *is a mapping that assigns to each variable* $x \in VAR$ *an element* $d \in D$ *such that* $v(x) \in OBJECT$*,* $v(a) \in PA$*, and* $v(i) \in AGENT$ *for any* $x \in RVAR, a \in AVAR$*, and* $i \in AGVAR$.

[1] since $\phi \boldsymbol{P}_i \phi \Rightarrow (\phi \wedge \neg\phi) \boldsymbol{P}_i (\phi \wedge \neg\phi) \Rightarrow \perp \boldsymbol{P}_i \perp \Rightarrow$ **False**.

Definition 11 (Valuation of terms) *For an ALX.3 model* $M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I\rangle$ *and a valuation of variables* v*, a valuation of terms* v_I *is a function that assigns to each term* $t \in TERM$ *an element in the domain* D*, which is defined as follows:*

$$t \in CON \Rightarrow v_I(t) = I_C(t);$$

$$t \in VAR \Rightarrow v_I(t) = v(t).$$

Definition 12 (Accessibility Relations for Actions) *We define an accessibility relation* $R^{a'}$ *in a model* $M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I\rangle$ *and a valuation* v *for each action* $a' \in ACTION$ *as follows.*

- $a \in ATERM, i \in AGTERM \Rightarrow R^{a_i} = \mathcal{R}(v_I(a), v_I(i))$,
- $a, b \in ACTION \Rightarrow R^{(a;b)} = R^a \circ R^b = \{\langle w, w'\rangle \in W \times W : (\exists w_1 \in W)(R^a w w_1 \text{ and } R^b w_1 w')\}$,
- $a, b \in ACTION \Rightarrow R^{(a \cup b)} = R^a \cup R^b$.

Definition 13 (Meaning function) *Let* FML *be as above and let* $M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I\rangle$ *be an ALX.3 model Let furthermore* v *be a valuation of variables in the domain* D*. Then the meaning function* $[\![\ ]\!]^v_M : FML \to \mathcal{P}(\mathcal{W})$ *is defined as follows:*

$$\begin{array}{lcl}
[\![p(t_1, ..., t_n)]\!]^v_M & = & \{w \in W : \langle v_I(t_1), v_I(t_2), ..., v_I(t_n)\rangle \in I_P(p, w)\} \quad \textit{where } p \in PRE_n. \\
[\![\neg\phi]\!]^v_M & = & W \setminus [\![\phi]\!]^v_M. \\
[\![\phi \wedge \psi]\!]^v_M & = & [\![\phi]\!]^v_M \cap [\![\psi]\!]^v_M. \\
[\![\exists x \phi]\!]^v_M & = & \{w \in W : (\exists d \in D)(w \in [\![\phi]\!]^{v(d/x)}_M)\}. \\
[\![\langle a\rangle\phi]\!]^v_M & = & \{w \in W : (\exists w' \in W)(R^a w w' \textit{ and } w' \in [\![\phi]\!]^v_M)\}. \\
[\![\phi \rightsquigarrow \psi]\!]^v_M & = & \{w \in W : cw(w, [\![\phi]\!]^v_M) \subseteq [\![\psi]\!]^v_M\}. \\
[\![\phi \boldsymbol{P}_i \psi]\!]^v_M & = & \{w \in W : cw(w, [\![\phi \wedge \neg\psi]\!]^v_M) \succ_{v_I(i)} cw(w, [\![\psi \wedge \neg\phi]\!]^v_M)\}. \\
[\![\boldsymbol{B}_i \phi]\!]^v_M & = & \{w \in W : (\forall w')(\langle w, w'\rangle \in \mathcal{B}_{v_I(i)} \Rightarrow w' \in [\![\phi]\!]^v_M)\}.
\end{array}$$

Definition 14 (The logic ALX.3) *Let* FML *be as above, let* Mod *be the class of all ALX.3 models, and let* $[\![\ ]\!]^v_M$ *be as above, defined for every model* $M \in Mod$*. We call the logic* $\langle FML, Mod, [\![\ ]\!]^v_M\rangle$ *ALX.3 logic.*
$\models$ *is defined as usual:*
Let $M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I\rangle$.

$$M \models \phi \overset{\text{def}}{\Longleftrightarrow} (\forall v \in V_D)(\forall w \in W)(M, w, v \Vdash \phi).$$

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

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

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

Definition 15 (ALX.3 inference system) *Let ALX3S be the following set of axioms and rules of inference.*

$(\boldsymbol{BA})$:	all tautologies of the first order logic.		
$(\boldsymbol{A1})$:	$\langle a\rangle\bot$	$\leftrightarrow$	$\bot$.
$(\boldsymbol{A2})$:	$\langle a\rangle(\phi\vee\psi)$	$\leftrightarrow$	$\langle a\rangle\phi \ \vee\langle a\rangle\psi$.
$(\boldsymbol{A3})$:	$\langle a;b\rangle\phi$	$\leftrightarrow$	$\langle a\rangle\langle b\rangle\phi$.
$(\boldsymbol{A4})$:	$\langle a\cup b\rangle\phi$	$\leftrightarrow$	$\langle a\rangle\phi\vee\langle b\rangle\phi$.
$(\boldsymbol{AU})$:	$[a]\forall x\phi$	$\leftrightarrow$	$\forall x[a]\phi$.
$(\boldsymbol{ID})$:	$\psi\rightsquigarrow\psi$.		
$(\boldsymbol{MPC})$:	$(\psi\rightsquigarrow\phi)$	$\rightarrow$	$(\psi\rightarrow\phi)$.
$(\boldsymbol{CC})$:	$(\psi\rightsquigarrow\phi)\wedge(\psi\rightsquigarrow\phi')$	$\rightarrow$	$(\psi\rightsquigarrow\phi\ wedge\phi')$.
$(\boldsymbol{MOD})$:	$(\neg\psi\rightsquigarrow\psi)$	$\rightarrow$	$(\phi\rightsquigarrow\psi)$.
$(\boldsymbol{CSO})$:	$[(\psi\rightsquigarrow\phi)\wedge(\phi\rightsquigarrow\psi)]$	$\rightarrow$	$[(\psi\rightsquigarrow\chi)\leftrightarrow(\phi\rightsquigarrow\chi)]$.
$(\boldsymbol{CV})$:	$[(\psi\rightsquigarrow\phi)\wedge\neg(\psi\rightsquigarrow\neg\chi)]$	$\rightarrow$	$[(\psi\wedge\chi)\rightsquigarrow\phi]$
$(\boldsymbol{CS})$:	$(\psi\wedge\phi)$	$\rightarrow$	$(\psi\rightsquigarrow\phi)$.
$(\boldsymbol{CEP})$:	$\phi\boldsymbol{P}_i\psi$	$\leftrightarrow$	$(\phi\wedge\neg\psi)\boldsymbol{P}_i(\neg\phi\wedge\psi)$.
$(\boldsymbol{N})$:	$\neg(\bot\boldsymbol{P}_i\phi),\neg(\phi\boldsymbol{P}_i\bot)$.		
$(\boldsymbol{TR})$:	$(\phi\boldsymbol{P}_i\psi)\wedge(\psi\boldsymbol{P}_i\chi)$	$\rightarrow$	$(\phi\boldsymbol{P}_i\chi.)$
$(\boldsymbol{PC})$:	$(\phi\boldsymbol{P}_i\psi)$	$\rightarrow$	$\neg((\phi\wedge\neg\psi)\rightsquigarrow\neg(\phi\wedge\neg\psi))\wedge$ $\neg((\psi\wedge\neg\phi)\rightsquigarrow\neg(\psi\wedge\neg\phi))$.
$(\boldsymbol{KB})$:	$\boldsymbol{B}_i\phi\wedge\boldsymbol{B}_i(\phi\rightarrow\psi)$	$\rightarrow$	$\boldsymbol{B}_i\psi$.
$(\boldsymbol{DB})$:	$\neg\boldsymbol{B}_i\bot$.		
$(4\boldsymbol{B})$:	$\boldsymbol{B}_i\phi$	$\rightarrow$	$\boldsymbol{B}_i\boldsymbol{B}_i\phi$.
$(\boldsymbol{BFB})$:	$\forall x\boldsymbol{B}_i\phi$	$\leftrightarrow$	$\boldsymbol{B}_i\forall x\phi$.
$(\boldsymbol{MP})$:	$\vdash\phi\ \&\ \vdash\phi\rightarrow\psi$	$\Rightarrow$	$\vdash\psi$.
$(\boldsymbol{G})$:	$\vdash\phi$	$\Rightarrow$	$\vdash\forall x\phi$.
$(\boldsymbol{NECA})$:	$\vdash\phi$	$\Rightarrow$	$\vdash[a]\phi$.
$(\boldsymbol{NECB})$:	$\vdash\phi$	$\Rightarrow$	$\vdash\boldsymbol{B}_i\phi$.
$(\boldsymbol{MONA})$:	$\vdash\langle a\rangle\phi\ \&\ \vdash\phi\rightarrow\psi$	$\Rightarrow$	$\vdash\langle a\rangle\psi$.
$(\boldsymbol{MONC})$:	$\vdash\phi\rightsquigarrow\psi\ \&\ \vdash\psi\rightarrow\psi'$	$\Rightarrow$	$\vdash\phi\rightsquigarrow\psi'$.
$(\boldsymbol{SUBA})$:	$\vdash(\phi\leftrightarrow\phi')$	$\Rightarrow$	$\vdash((\langle a\rangle\phi)\leftrightarrow((\langle a\rangle\phi')$.
$(\boldsymbol{SUBC})$:	$\vdash(\phi\leftrightarrow\phi')\ \&\ \vdash(\psi\leftrightarrow\psi')$	$\Rightarrow$	$\vdash(\phi\rightsquigarrow\psi)\leftrightarrow(\phi'\rightsquigarrow\psi')$.
$(\boldsymbol{SUBP})$:	$\vdash(\phi\leftrightarrow\phi')\ \&\ \vdash(\psi\leftrightarrow\psi')$	$\Rightarrow$	$\vdash(\phi\boldsymbol{P}_i\psi)\leftrightarrow(\phi'\boldsymbol{P}_i\psi')$.

Most axioms are straightforward. As usual, we have the tautologies (BA). Since ALX.3 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 (A2) (they also distribute over conjunction in one direction, but the corresponding axiom is redundant). (A3) characterizes the sequencing operator ';' and (A4) does the same for the indeterminate choice of actions. (AU) establishes the Barcan formula for universal action modalities. We have the Barcan formula because the underlying domain D is the same in all possible worlds.

The next seven axioms characterize the intensional conditional. Informally speaking, they specify syntactically the meaning of "ceteris paribus" in ALX.3. They are fairly standard, and, with the exception of (CC), they already provide a characterization of Lewis' system **VC**, which, in turn, is an adaptation of Stalnaker's conditional logic for non-unique closest worlds. (ID) establishes the triviality that ψ is true in all closest ψ worlds; (MPC) relates the intensional and the material conditional in the obvious way: so if ϕ would hold given ψ, then, if ψ actually does hold, ϕ must also hold. Conjunction distributes over the causality operator in one way (CC). (MOD) rules out the eventuality of closest absurd worlds; (CSO) gives an identity condition for closest worlds, (CV) establishes a cautious monotony for the intensional conditional, and (CS) relates the conjunction to the intensional conditional. Replacing (CS) by

$$(\phi \rightsquigarrow \psi) \vee (\phi \rightsquigarrow \neg\psi)$$

would return Stalnaker's original systems, as the new axiom would require the uniqueness of the closest possible world.

The next four axioms characterize the preference relation. (CEP) states the conjunction expansion principle. (IRE) confirms the irreflexivity of the $\boldsymbol{P}_i$ operator. (N) establishes "normality" and (TR) transitivity. As noted before, (TR) would go if its semantic equivalent, (TRAN), goes, so we could have non-transitive preferences. The axiom (PC) says that if an agent i prefers ϕ to ψ, then both $\phi \wedge \neg\psi$ and $\psi \wedge \neg\phi$ are possible.

The last four axioms characterize the belief operator. As pointed out above, our belief operator is designed to represent subjective knowledge. (KB) is standard in epistemic logic, but it is often criticised, since it requires logical omniscience with respect to the material conditional. On the other hand, one would expect rational agents to draw correct logical inferences, when necessary, so not having (KB) might be worse. (DB) rules out the belief in absurdities, (4B) establishes positive self-introspection for beliefs, and (BFB) is the Barcan formula for beliefs. These four axioms give a standard characterization of subjective knowledge. Together with the inference rules (MP) and (NECB), they turn the belief operation into a weak S4 system.

The remaining expressions characterize ALX.3's inference rules.

We have modus ponens and generalization for obvious reasons. By the same token, we have the necessitation rule for the universal action modality: if indeed, ϕ is true in all worlds, then all activities will lead to ϕ-worlds; by the same token, we have the necessitation rule for beliefs. (MONA) connects the meaning of the action modality with the meaning of the material conditional. We have right monotonicity for the intensional conditional but *not* left monotonicity, Furthermore, logically equivalent propositions are substitutable in action- conditional- and preference formulae (SUBA), (SUBC), (SUBP). Note that we are *not* having monotonicity for preferences. Because of this, we are able to avoid the counterintuitive deductive closure of goals that marrs other action logics.

6 Appendix 2: Some Proofs

Proposition 3 *The deontic ALX.3 logic can avoid (i) Ross's paradox, (ii) Penitent's paradox, and (iii) Good Samaritan paradox.*

PROOF: (i) In order to prove that the deontic logic in ALX.3 can avoid Ross's paradoxes, we have to show that $\mathbf{O_i} \wedge \neg\mathbf{O_i}(\phi \vee \psi)$ is satisfiable.

$$
\begin{array}{lll}
 & \mathbf{O_i}\phi \wedge \neg\mathbf{O_i}(\phi \vee \psi) \text{ is satisfiable} & \\
\Leftrightarrow & \mathbf{F_i}\neg\phi \wedge \neg\mathbf{F_i}\neg(\phi \vee \psi) \text{ is satisfiable} & (\text{Definition of } \mathbf{O_i}) \\
\Leftrightarrow & \mathbf{F_i}\neg\phi \wedge \neg\mathbf{F_i}(\neg\phi \wedge \neg\psi) \text{ is satisfiable} & (\text{Meta reasoning}) \\
\Leftrightarrow & \mathbf{Bad}_{sg}\mathbf{Good}\neg\phi \wedge \neg\mathbf{Bad}_{sg}\mathbf{Good}(\neg\phi \wedge \neg\psi) \text{ is satisfiable} & (\text{Definition of } \mathbf{F_i}) \\
\Leftrightarrow & \neg(\mathbf{Good}\neg\phi)\boldsymbol{P}_{sg}(\mathbf{Good}\neg\phi)\wedge & \\
 & \neg((\neg\mathbf{Good}(\neg\phi \wedge \neg\psi))\boldsymbol{P}_{sg}\mathbf{Good}(\neg\phi \wedge \neg\psi)) \;\; \text{is satisfiable} & (\text{Definition of } \mathbf{Bad})
\end{array}
$$

In the following, we will construct a model M such that there exists a world which

$$M, w, v \Vdash \neg(\mathbf{Good}\neg p)\boldsymbol{P}_{sg}(\mathbf{Good}\neg p) \wedge \neg((\neg\mathbf{Good}(\neg p \wedge \neg q))\boldsymbol{P}_{sg}\mathbf{Good}(\neg p \wedge \neg q)).$$

The model is constructed as follows:

$M = \langle O, PA, AGENT, W, cw, \succ, \mathcal{R}, \mathcal{B}, I\rangle$,
where
$O = \emptyset$,
$PA = \emptyset$,
$AGENT = \{sg, ag\}$,
$W = \{w_0, w_1, w_2, w_3\}$,
cw is defined by the minimal difference between two worlds,
$\succ (sg) = \{\langle\{w_1, w_2\}, \{w_0\}\rangle\}$;
$\succ (ag) = \{\langle\{w_2\}, \{w_0\}\rangle\}$,
$\mathcal{R} = \emptyset$,
$\mathcal{B} = \emptyset$,
$I_P(p, w_0) = \{\langle ag\rangle\}$,
$I_P(p, w_1) = \{\langle ag\rangle\}$;
$I_p(q, w_0) = \{\langle ag\rangle\}$,
$I_p(q, w_2) = \{\langle ag\rangle\}$.

Let v be a variable evaluation $v(i) = ag$.

So, we have $[\![\neg p(i)\boldsymbol{P}_i p(i)]\!]^v_M = \{w_0\}$; $[\![\neg(\neg p(i)\boldsymbol{P}_i p(i))]\!]^v_M = \{w_1, w_2, w_3\}$. Furthermore, $[\![(\neg p(i) \wedge \neg q(i))\boldsymbol{P}_i\neg(\neg p(i) \wedge \neg q(i))]\!]^v_M = \emptyset = [\![\bot]\!]^v_M$.

Therefore, $cw(w_0, [\![\neg(\neg p(i)\boldsymbol{P}_i p(i))]\!]^v_M) = \{w_1, w_2\}$. By the truth condition, we have

$$M, w_0, v \Vdash \neg(\mathbf{Good}\neg p(i))\boldsymbol{P}_{sg}(\mathbf{Good}\neg p(i)).$$

Furthermore, by the normality, we have

$$M, w_0, v \Vdash \neg((\neg\mathbf{Good}(\neg p(i) \wedge \neg q(i)))\boldsymbol{P}_{sg}\mathbf{Good}(\neg p(i) \wedge \neg q(i))).$$

So,

$$\neg(\mathbf{Good}\neg p(i))\boldsymbol{P}_{sg}(\mathbf{Good}\neg p(i)) \wedge \neg((\neg\mathbf{Good}(\neg p(i) \wedge \neg q(i)))\boldsymbol{P}_{sg}\mathbf{Good}(\neg p(i) \wedge \neg q(i)))$$

is satisfiable. That completes the proof for (i).

For (ii) and (iii), we follow the same type of the proof. We do not go into the details here. □

References

[1] Anderson, A. R., A reduction of deontic logic to alethic modal logic. *Mind* 67, 1958, 100-103.

[2] Chisholm, R., and Sosa, E., On the logic of "intrinsically better", *American Philosophical Quarterly* **3** (1966) 244-249.

[3] Danielsson, S., *Preference and obligation*, (Filosofischa föreningen, Uppsala, 1968).

[4] Harel, D.. Dynamic logic, in: D. Gabbay and F. Guenthner, eds., *Handbook of Philosophical Logic*, Vol.II, (D. Reidel Publishing Company, 1984) 497-604.

[5] Huang, Z., Masuch, M., and Pólos, L., ALX, an action logic for agents with bounded rationality, *Artificial Intelligence* 82 (1996) 100-153.

[6] Huang, Z., Masuch, M., and Pólos, L., ALX2, a quantifier ALX logic, CCSOM Research Report 93-99, (1993).

[7] Huang, Z., *Logics for Agents with Bounded Rationality*, ILLC Dissertation Series 1994-10.

[8] Huang, Z., and Masuch, M., ALX.3: a multi-agent action logic, *the Proceedings of AAAI95 Spring Symposium*, (1995).

[9] Meyer, J-J., A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic, *Notre Dame J. of Formal Logic* 29 (1988) 109-136.

[10] Meyer, J-J., and Wieringa, R. J., Deontic logic: a concise overview, in Meyer, J.-J., and Wieringa, R. J., (eds.) *Deontic Logic in Computer Science*, (John Wiley & Sons, 1993).

[11] Stalnaker, R., A theory of conditionals in: *Studies in Logical Theory, American Philosophical Quarterly* **2** (1968) 98-122

[12] von Wright, G. H., *The Logic of Preference*, (Edinburgh, 1963).

[13] von Wright, G. H., The logic of preference reconsidered, *Theory and Decision* **3** (1972) 140-169.