A logic of higher-order preferences

Junli Jiang[1] ⬛ · **Pavel Naumov[2]** ⬛

Abstract
If an agent prefers one kind of agents to the other agents, then the agent has first-order preferences. If the agent prefers agents with one kind of preferences to the other agents, then the agent has second-order preferences. The article proposes a sound, complete, and decidable logical system capable of expressing higher-order preferences.

Keywords Egocentric · Higher-order preferences · Axiomatisation · Completeness

1 Introduction

Logics for reasoning about preferences have been the subject of studies in philosophy, logic, and artificial intelligence for a long time. As with many other modal logics, the focus of the initial works was on finding the right language and writing intuitively acceptable axioms for this language. In the case of preferences, this approach was explored by Halldén (1957), Åqvist (1962), Chisholm and Sosa (1966), and Von Wright (1963). After possible world semantics for modal logics became widely accepted, it became common to start not with intuition, but with a formal semantics of the proposed modality. Following this tradition, van Benthem, Girard, and Roy (2009) consider the modality $[>]\varphi$ meaning "statement φ holds in all worlds better than the current one". The same modality is also investigated in (Liu, 2011) and (Christoff et al., 2021). Even earlier, Lang, van der Torre, and Weydert (2002) considered a conditional version of this modality.

✉ Pavel Naumov
  p.naumov@soton.ac.uk

  Junli Jiang
  walk08@swu.edu.cn

[1] Institute of Logic and Intelligence, Southwest University, Chongqing, China

[2] University of Southampton, Southampton, UK

◇ Springer

In addition to modality $[>]\varphi$, one can also consider modality $[\not>]\varphi$, meaning "statement φ holds in all worlds not better than the current one".[1] Then, formula $[\not>]\neg\varphi$ means that all worlds in which φ holds are better than the current world. One can combine modalities $[>]$ and $[\not>]$ to express the fact that statement φ holds in worlds better than current and only them: $[>]\varphi \wedge [\not>]\neg\varphi$.

In this article, we study preferences between agents rather than between worlds. Modalities $[>]$ and $[\not>]$ can be adopted for this purpose by using an egocentric logical system (Prior, 1968). In traditional logics, formulae capture properties of possible worlds. In egocentric logics, they capture the properties of agents. As a result, the semantics of an egocentric logic can be defined in terms of a satisfaction relation $a \Vdash \varphi$ between an agent a and a formula φ rather than a relation $w \Vdash \varphi$ between a world w and a formula φ. In such a setting, statement $a \Vdash [>]\varphi$ means that property φ holds for all agents better than agent a. Similarly, $a \Vdash [\not>]\varphi$ means that φ is true for all agents who are not better than agent a.

Note that modalities $[>]$ and $[\not>]$ capture *global* preferences on agents, not the preferences of *individual* agents. The goal of the current work is to study individual preferences of agents about groups of agents. More specifically, we study the relation "agent a prefers any agent with property φ to any agent who does not have this property". If the relation holds, then we write $a \Vdash \mathsf{L}\varphi$ and say that agent a "likes" φ. Thus, we use the word "preference" to refer to a relation on agents and the word "likes" to refer to the modality that captures the properties of the preferences. Modality L can be nested to express *higher-order preferences*. For example, the statement $a \Vdash \mathsf{LL}$"is a philosopher" means that agent a likes those who like philosophers. Statement $a \Vdash \neg\mathsf{L}\neg\mathsf{L}$"is a philosopher" means that agent a does not like those who do not like philosophers.

In some situations, there is no need for the "likes" modality L because in those situations this modality can be expressed through already studied modality F and its variation NF. Modality "for all friends" F was proposed by Seligman, Liu, and Girard (2011) and is also used in (Seligman et al., 2013), (Christoff and Hansen, 2015), and (Christoff et al., 2016). It presupposes that each agent partitions the set of all agents into "friends" and "non-friends". The statement $a \Vdash \mathsf{F}\varphi$ is true if all friends of agent a have property φ. One can similarly consider the "for all non-friends" modality NF. The statement $a \Vdash \mathsf{NF}\varphi$ is true if all non-friends of agent a have property φ. Let us now go back to preferences and consider an agent a with "good-or-bad" preferences in which agent a divides all agents into "good" and "bad". Agent a prefers each good agent to each bad agent and has no other preferences. In such a situation, we can call "good" agents to be a's friends and "bad" agents to be a's non-friends. Then, the formula $\mathsf{L}\varphi$ is equivalent to the formula $\mathsf{F}\varphi \wedge \mathsf{NF}\neg\varphi$.

To further illustrate "good-or-bad" preferences, let us consider an example inspired by the quote

"Steve Jobs has a saying that *A* players hire *A* players; *B* players hire *C* players; and *C* players hire *D* players." (Kawasaki, 2004, p. 101)

---

[1] We do not assume that the order is total.

Fig. 1 The world according to
Steve Jobs
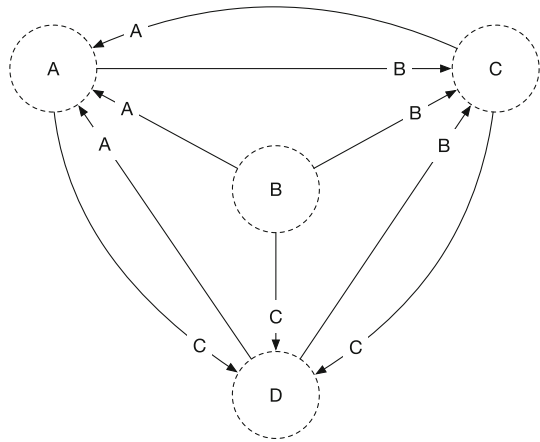
We capture this quote by the diagram depicted in Fig. 1. Here all agents are partitioned into groups A, B, C, and D. The hiring preference relation of each group is depicted in the figure using directed edges. We assume that D-players have no preferences. Note that, in this example, each agent $c \in C$ prefer any D-player to any non-D-player. Thus, by our definition of the "likes" modality,

$$c \Vdash \mathsf{L}\text{"is a } D \text{ player"} \qquad \forall c \in C. \tag{1}$$

Observe from Fig. 1 that the same is not true[2] for agents who are not C-players:

$$x \nVdash \mathsf{L}\text{"is a } D \text{ player"} \qquad \forall x \in A \cup B \cup D. \tag{2}$$

Note that any agent $b \in B$ prefers any agent in group C to any agent in group $A \cup B \cup D$, see Fig. 1. Thus, taking into account statements (1) and (2), any agent $b \in B$ prefers any agent for whom L"is a D player" is true to any agent for whom L"is a D player" is false. Thus, any agent $b \in B$ likes those for whom L"is a D player" is true:

$$b \Vdash \mathsf{LL}\text{"is a } D \text{ player"} \qquad \forall b \in B.$$

In other words, B-players hire those who hire D-players.

Note that although, overall, the structure of the preferences in the Steve Jobs example is not so simple, the preferences of each individual agent are still of "good-or-bad" type. For example, for any agent $b \in B$, the "good" group is C and the "bad" group is $A \cup B \cup D$. Thus, "likes" modality L still can be expressed through modalities F and NF.

The situation is different for non "good-or-bad" preferences. For example, consider an agent a who divides all agents into three disjoint groups: "very good", "good", and "bad". Suppose that agent a prefers each "very good" agent to each "good" and each

---

[2] We assume that sets D and $A \cup B \cup C$ are not empty.

"bad" agent. Also, *a* prefers each "good" agent to each "bad" agent. Then, *a* likes all agents in the first group:

$$a \Vdash \mathsf{L}\text{"is a 'very good' agent"}$$

because *a* prefers each "very good" agent to any agent in the other two groups. At the same time, agent *a* also likes all agents in the first two groups:

$$a \Vdash \mathsf{L}\text{"is not a 'bad' agent"}$$

becuse agent *a* prefers each agent who is not "bad" to any agent who is "bad". Next, we will observe that in this setting modality $\mathsf{L}\varphi$ is not equivalent to the formula $\mathsf{F}\varphi \wedge \mathsf{NF}\neg\varphi$ no matter how we define who are the friends and the non-friends of agent *a*. Indeed, no matter how we divide agents into friends and non-friends, there is no way to satisfy both of the following formulae:

$$a \Vdash \mathsf{F}\text{"is a 'very good' agent"} \wedge \mathsf{NF}\neg\text{"is a 'very good' agent"} \tag{3}$$

and

$$a \Vdash \mathsf{F}\text{"is not a 'bad' agent"} \wedge \mathsf{NF}\neg\text{"is not a 'bad' agent"}. \tag{4}$$

For example, if only "very good" agents are identified as friends, then statement (3) is true, but statement (4) is not. At the same time, if "very good" and "good" agents are identified as friends, then it is the other way around.

In this article, we propose a formal semantics for modality L in the egocentric setting and give a sound, complete, and decidable axiomatization of the properties of this modality. The completeness theorem is using a new technique based on relation $\sqsubseteq$, see Sect. 6.3.

The rest of this article is organised as follows. In Sect. 2, we give a formal definition of a preference model that we use later to give the semantics of modality L. In Sect. 3 we describe the syntax and semantics of our language. Section 4 lists the axioms and inference rules of our system. We prove its soundness and completeness in Sect. 5 and Sect. 6, respectively. In Sect. 7, we define "conditional likes" modality, list some of the axioms for it and prove that it is not definable through modality L. Section 8 concludes. A preliminary version of this work, containing a model checking algorithm for a richer language, but without the proof of completeness and the discussion of the conditional likes modality, appeared as (Jiang and Naumov, 2022).

2 Preference model

In this section, we introduce the notion of a preference model that serves as a foundation for the semantics of our logical system. In the article, we assume a fixed set of propositional variables $\mathcal{P}$.

Definition 1 A tuple $(\mathcal{A}, \{\prec_a\}_{a \in \mathcal{A}}, \pi)$ is called a preference model if

1. $\mathcal{A}$ is a (possibly empty) finite set of "agents",
2. $\prec_a$ is a strict partial order "preference" relation on $\mathcal{A}$, for each agent $a \in \mathcal{A}$,
3. $\pi(p)$ is a subset of $\mathcal{A}$ for each variable $p \in \mathcal{P}$.

Note that, in general, strict and non-strict order relations are definable through each other: $x \preceq y \equiv (x \prec y \lor x = y)$ and $x \prec y \equiv (x \preceq y \land x \neq y)$. Thus, it does not matter which of them is chosen as the primitive relation and which is defined through the primitive. Hence, it is not significant that in the above definition we have chosen the strict relation $\prec_a$ as our primitive.

Note that propositional variables, just like all formulae in our language, represent noun-free fragments of propositions such as "is a philosopher". Prior calls them *subject-less predicates* (1968). Grove and Halpern use the term *relative propositions* because they are relative to agents (Grove and Halpern, 1991; Grove & Halpern, 1993; Grove, 1995). To reflect this, item 3 of Definition 1 specifies the valuation $\pi(p)$ of a proposition variable p as a set of agents. Informally, these are the agents for which the propositional variable p is true.

3 Syntax and semantics

The language Φ of our system is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{L}\varphi,$$

where $p \in \mathcal{P}$ is a propositional variable. We read L as "likes those who". We assume that disjunction $\lor$, conjunction $\land$, and biconditional $\leftrightarrow$ are defined in our language in the standard way. Next, we define the semantics of our logical system.

Definition 2 For any formula $\varphi \in \Phi$ and any agent $a \in \mathcal{A}$ of a preference model $(\mathcal{A}, \{\prec_a\}_{a \in \mathcal{A}}, \pi)$, the satisfaction relation $a \Vdash \varphi$ is defined recursively as follows.

1. $a \Vdash p$, if $a \in \pi(p)$,
2. $a \Vdash \neg\varphi$, if $a \nVdash \varphi$,
3. $a \Vdash \varphi \to \psi$, if $a \nVdash \varphi$ or $a \Vdash \psi$,
4. $a \Vdash \mathsf{L}\varphi$, when for all agents $b, c \in \mathcal{A}$, if $b \nVdash \varphi$ and $c \Vdash \varphi$, then $b \prec_a c$.

As we have discussed in Sect. 2, the choice to use strict preference relation $\prec_a$ in Definition 1 is not important. Irrespective of the choice of the primitive relation in Definition 1, one can consider either strict or non-strict preference relation in item 4 of Definition 2. We have chosen to use the strict preference relation there. The non-strict version of the same item will look like:

$$a \Vdash \mathsf{L}\varphi, \quad \text{when for all agents } b, c \in \mathcal{A}, \text{ if } b \nVdash \varphi \text{ and } c \Vdash \varphi, \text{ then } b \preceq_a c. \quad (5)$$

However, note that the assumptions $b \nVdash \varphi$ and $c \Vdash \varphi$ in the above statement imply that $b \neq c$. Thus, statement (5) is equivalent to item 4 of Definition 2 and it does not

matter which of them is used in Definition 2. To summarise, although we have chosen to use strict order in Definition 1 and Definition 2, nothing in our results will change if either of those places (or even in both of them) non-strict order is used instead of the strict one.

Some of the existing works on logics of preferences are based on *ceteris paribus* principle (Von Wright, 1963). If this principle is applied to modality "likes", then one would say that an agent a likes φ if she prefers agents for whom φ is true to those for who φ is not true, *everything else being equal*. Capturing the "everything else being equal" assumption semantically is a non-trivial task. One of the possible ways to do it is proposed in (Van Benthem et al., 2009). In this article, we do *not* use ceteris paribus principle[3]. Item 4 of the above definition states that an agent a likes φ if she prefers *each* agent for whom φ is true to *each* agent for whom φ is not true. This definition is similar to how *strong belief* is defined in the literature: an agent strongly believes in φ if she finds all epistemically possible worlds in which φ is true to be more plausible than those where φ is false (Baltag and Smets, 2009; Lorini, 2021). Lorini also proposes to define *strong desires* in a similar way.

We use $A\varphi$ as an abbreviation for the formula $\varphi \wedge L\varphi \wedge L\neg\varphi$. Because of the next lemma, we read A as "for all agents".

Lemma 1 *$a \Vdash A\varphi$ iff $b \Vdash \varphi$ for all agents $b \in \mathcal{A}$.*

Proof ($\Rightarrow$) : Suppose that there is an agent $b \in \mathcal{A}$ such that $b \nVdash \varphi$. Note that the assumption $a \Vdash A\varphi$ of the lemma implies that $a \Vdash \varphi$ and $a \Vdash L\varphi$. Then, by item 4 of Definition 2, the assumption $b \nVdash \varphi$ implies that $b \prec_a a$.

At the same time, the assumption $b \nVdash \varphi$ implies $b \Vdash \neg\varphi$ by item 2 of Definition 2. Note that the assumption $a \Vdash A\varphi$ of the lemma also implies that $a \Vdash \varphi$ and $a \Vdash L\neg\varphi$. Then, $a \prec_a b$ again by item 4 of Definition 2.

Finally, note that statements $b \prec_a a$ and $a \prec_a b$ are inconsistent because relation $\prec_a$ is a strict partial order by item 2 of Definition 1.
($\Leftarrow$) : Suppose that $a \nVdash \varphi \wedge L\varphi \wedge L\neg\varphi$. Thus, one of the following cases takes place.
Case I: $a \nVdash \varphi$. Then, there is an agent b such that $b \nVdash \varphi$.
Case II: $a \nVdash L\varphi$. Hence, by item 4 of Definition 2, there are agents $b, c \in \mathcal{A}$ such that $b \nVdash \varphi$, $c \Vdash \varphi$, and $b \nprec_a c$. So, there is an agent b such that $b \nVdash \varphi$.
Case III: $a \nVdash L\neg\varphi$. Thus, again by item 4 of Definition 2, there are agents $b, c \in \mathcal{A}$ such that $b \nVdash \neg\varphi$, $c \Vdash \neg\varphi$, and $b \nprec_a c$. Therefore, $c \nVdash \varphi$ by item 2 of Definition 2. □

The lemma below follows from item 4 of Definition 2.

Lemma 2 *For any preference model, if $a \Vdash \varphi \leftrightarrow \psi$ for each agent $a \in \mathcal{A}$, then $a \Vdash L\varphi \leftrightarrow L\psi$ for each agent $a \in \mathcal{A}$.*

4 Axioms

In this section, we list the axioms and the inference rules of our logical system. In addition to tautologies in language Φ, the system contains the following axioms.

---

[3] In Sect. 7, we discuss conditional like modality which, partially, captures the idea behind ceteris paribus principle.

1. Reflexivity: $\mathsf{A}\varphi \rightarrow \varphi$,
2. Distributivity: $\mathsf{A}(\varphi \rightarrow \psi) \rightarrow (\mathsf{A}\varphi \rightarrow \mathsf{A}\psi)$,
3. Euclidianity: $\neg\mathsf{A}\varphi \rightarrow \mathsf{A}\neg\mathsf{A}\varphi$,
4. Substitution: $\mathsf{A}(\varphi \leftrightarrow \psi) \rightarrow (\mathsf{L}\varphi \rightarrow \mathsf{L}\psi)$,
5. Coherence of Preferences: $\mathsf{L}\varphi \wedge \mathsf{L}\psi \rightarrow \mathsf{A}(\varphi \rightarrow \psi) \vee \mathsf{A}(\psi \rightarrow \varphi)$.

The meaning of the first four axioms above is straightforward. The Coherence of Preferences axiom states that if an agent likes those for whom φ is true and also those for whom ψ is true, then either each φ-agent is a ψ-agent or each ψ-agent is a φ-agent. This is the most non-trivial axiom of our system. We prove its soundness in Lemma 10.

We write $\vdash \varphi$ and say that formula $\varphi \in \Phi$ is a *theorem* of our logical system if it can be derived from the above axioms using the Modus Ponens and the Necessitation inference rules:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi} \qquad \frac{\varphi}{\mathsf{A}\varphi}.$$

In addition to unary relation $\vdash \varphi$ we also use binary relation $X \vdash \varphi$. We write $X \vdash \varphi$ if formula φ is derivable from the *theorems* of our logical system and the set of additional formulae X using *only* the Modus Ponens inference rule. It is easy to see that statements $\varnothing \vdash \varphi$ and $\vdash \varphi$ are equivalent. We say that set X is consistent if there is no formula $\varphi \in \Phi$ such that $X \vdash \varphi$ and $X \vdash \neg\varphi$.

Lemma 3 (Lindenbaum) *For any set of formulae X, any consistent subset of X can be extended to a maximal consistent subset of X.*

Proof The standard proof of Lindenbaum's lemma ((Mendelson, 2009), Proposition 2.14) applies here. □

The proofs of the next three standard lemmas can be found in the appendix.

Lemma 4 (deduction) *If $X, \varphi \vdash \psi$, then $X \vdash \varphi \rightarrow \psi$.*

Lemma 5 (Transitivity) $\vdash \mathsf{A}\varphi \rightarrow \mathsf{AA}\varphi$.

Lemma 6 *If $\varphi_1, \ldots, \varphi_n \vdash \psi$, then $\mathsf{A}\varphi_1, \ldots, \mathsf{A}\varphi_n \vdash \mathsf{A}\psi$.*

5 Soundness

In this section, we prove the soundness of our logical system.

Theorem 1 (soundness) *If $\vdash \varphi$, then $a \Vdash \varphi$ for each agent a of each preference model.*

The soundness of the Reflexivity axiom as well as of the Modus Ponens and the Necessitation inference rules is straightforward. To improve readability, we prove the soundness of each remaining axiom as a separate lemma.

Lemma 7 (Distributivity) *If $a \Vdash \mathsf{A}(\varphi \rightarrow \psi)$ and $a \Vdash \mathsf{A}\varphi$, then $a \Vdash \mathsf{A}\psi$.*

Proof Consider any agent $b \in \mathcal{A}$. By Lemma 1, it suffices to show that $b \Vdash \psi$. Indeed, by the same Lemma 1, the assumptions $a \Vdash A(\varphi \to \psi)$ and $a \Vdash A\varphi$ imply $b \Vdash \varphi \to \psi$ and $b \Vdash \varphi$, respectively. Therefore, $b \Vdash \psi$ by item 3 of Definition 2. ☐

Lemma 8 (Euclidianity) *If $a \nVdash A\varphi$, then $a \Vdash A\neg A\varphi$.*

Proof Consider any $b \in \mathcal{A}$. By Lemma 1, it suffices to show that $b \Vdash \neg A\varphi$.

Again by Lemma 1, the assumption $a \nVdash A\varphi$ implies that there is an agent $c \in \mathcal{A}$ such that $c \nVdash \varphi$. Thus, $b \nVdash A\varphi$ also by Lemma 1. Therefore, $b \Vdash \neg A\varphi$ by item 2 of Definition 2. ☐

Lemma 9 (Substitution) *If $a \Vdash A(\varphi \leftrightarrow \psi)$ and $a \Vdash L\varphi$, then $a \Vdash L\psi$.*

Proof Consider any agents $b, c \in \mathcal{A}$ such that $b \nVdash \psi$ and $c \Vdash \psi$. By item 4 of Definition 2, it suffices to show that $b \prec_a c$.

By Lemma 1, the assumption $a \Vdash A(\varphi \leftrightarrow \psi)$ implies that $b \Vdash \varphi \leftrightarrow \psi$ and $c \Vdash \varphi \leftrightarrow \psi$. Hence, $b \nVdash \varphi$ and $c \Vdash \varphi$ by the assumptions $b \nVdash \psi$ and $c \Vdash \psi$, respectively. Therefore, $b \prec_a c$ by item 4 of Definition 2 and the assumption $a \Vdash L\varphi$ of the lemma. ☐

Lemma 10 (Coherence of preferences) *If $a \Vdash L\varphi$ and $a \Vdash L\psi$, then either $a \Vdash A(\varphi \to \psi)$ or $a \Vdash A(\psi \to \varphi)$.*

Proof Suppose that $a \nVdash A(\varphi \to \psi)$ and $a \nVdash A(\psi \to \varphi)$. Thus, by Lemma 1, there are agents $b, c \in \mathcal{A}$ such that $b \nVdash \varphi \to \psi$ and $c \nVdash \psi \to \varphi$. Hence, by item 3 of Definition 2,

$$b \Vdash \varphi, \qquad b \nVdash \psi, \qquad c \Vdash \psi, \qquad c \nVdash \varphi. \qquad (6)$$

By item 4 of Definition 2, the assumption $a \Vdash L\varphi$ of the lemma and parts $c \nVdash \varphi$ and $b \Vdash \varphi$ of statement (6) imply

$$c \prec_a b. \qquad (7)$$

At the same time, by item 4 of Definition 2, the assumption $a \Vdash L\psi$ of the lemma and parts $c \Vdash \psi$ and $b \nVdash \psi$ of statement (6) imply $b \prec_a c$. Hence, $c \nprec_a b$ because $\prec_a$ is a strict partial order relation by Definition 1. This is a contradiction with statement (7). ☐

6 Completeness and decidability

In this section, we prove the completeness of our logical system and also show that our system is decidable. As usual, the proof of completeness relies on a construction of a canonical model. In modal logic, the worlds of a canonical model are usually defined as maximal consistent sets of formulae. Since the formal semantics of our language, given in Definition 2, defines the satisfaction $\Vdash$ as a relation between an agent and a formula, in our system, the agents play a role somewhat similar to the role of worlds

in traditional modal logics. In particular, we define agents as maximal consistent sets of formulae.

There are two major challenges that we faced while constructing the canonical model. First, Definition 1 requires the set of agents $\mathcal{A}$ to be finite. To guarantee the finiteness of the canonical model, we employ the filtration technique. Usually, this technique requires to restrict the language to subformulae of the formulae φ for which the canonical model is constructed. This approach fails in our case because the proof of completeness requires us to deal with formulae which are not subformulae of formula φ. To solve the problem, instead of subformulae of φ we consider a larger "saturated" set of formulae. We define the notion of a saturated set and prove its property in Sect. 6.1.

The other challenge in the proof of completeness is to give a definition of the preference relation $\prec_a$ that guarantees that the relation is transitive. We give such a definition and a non-trivial proof of transitivity in Sect. 6.2 as a part of the canonical model description.

At the core of our proof of completeness is the truth lemma proven by induction on the size of a formula. In our case, this lemma states that, in the canonical model, $a \Vdash \psi$ if and only if formula ψ belongs to the maximal consistent set a of formulae. To make the proof of this lemma more readable, we separated the major case in the induction step of the proof into a separate lemma stated and proven in Sect. 6.3. We use this lemma in Sect. 6.4 to prove the truth lemma and to finish the proof of completeness.

6.1 Saturated set

As a part of the proof, we use the technique known as filtration to guarantee that the set of agents in the canonical model is finite. To do filtration, we restrict some formulae to those in a finite "saturated" set Φ_0. Notions similar to our "saturated" set have been used before (Goranko & van Drimmele, 2006, p. 102).

Definition 3 Set of formulae $\Phi_0 \subseteq \Phi$ is saturated if it is closed with respect to subformulae and

1. if $\varphi \in \Phi_0$ and formula φ is not a negation, then $\neg\varphi \in \Phi_0$,
2. if $\mathsf{A}\varphi \in \Phi_0$, then $\neg\varphi \in \Phi_0$,
3. if $\mathsf{L}\varphi \in \Phi_0$, then $\neg\varphi \in \Phi_0$.

Lemma 11 *Any finite set of formulae can be extended to a finite saturated set of formulae.*

Proof Let $X \subseteq \Phi$ be any finite set of formulae. Consider the set

$$\Phi_0 = \{\psi, \neg\psi \mid \psi \text{ is a subformula of some formula in set } X\}.$$

Note that set Φ_0 is a finite extension of set X. To finish the proof of the lemma, it suffices to verify conditions 1–3 from Definition 3.

To verify condition 1, consider any formula $\varphi \in \Phi_0$ which is not a negation. Thus, by the choice of set Φ_0, formula φ itself must be a subformula of some formula in set X. Therefore, $\neg\varphi \in \Phi_0$ by the choice of set Φ_0.

To verify condition 3, consider any formula $L\varphi \in \Phi_0$. Again by the choice of set Φ_0, formula $L\varphi$ must be a subformula of some formula in set X. Then, formula φ is also a subformula of some formula in set X. Therefore, $\neg\varphi \in \Phi_0$ by the choice of set Φ_0.

To verify condition 2, consider any formula $A\varphi \in \Phi_0$. Thus, by the definition of modality A, it follows that $\varphi \wedge L\varphi \wedge L\neg\varphi \in \Phi_0$. Then, again by the choice of set Φ_0, formula φ must be a subformula of some formula in set X. Hence, $\neg\varphi \in \Phi_0$ by the choice of set Φ_0. $\qquad\square$

6.2 Canonical preference model

In this subsection, we construct a canonical preference model for our logical system. First, for any saturated set of formulae Φ_0, by set Φ_0^* we denote the set of formulae

$$\Phi_0^* = \Phi_0 \cup \{A\varphi \mid \varphi \in \Phi_0\} \cup \{A\neg\varphi \mid \varphi \in \Phi_0\}$$
$$\cup \{A(\varphi \to \psi) \mid \varphi, \psi \in \Phi_0\} \cup \{\neg A(\varphi \to \psi) \mid \varphi, \psi \in \Phi_0\}. \qquad (8)$$

Next, for any saturated set of formulae Φ_0 and any maximal consistent set $a_0 \subseteq \Phi_0^*$, we define canonical preference model

$$N(\Phi_0, a_0) = (\mathcal{A}, \{\prec_a\}_{a \in \mathcal{A}}, \pi).$$

Definition 4 $\mathcal{A}$ is the set of all maximal consistent sets $a \subseteq \Phi_0^*$ such that

$$\{A\varphi \mid A\varphi \in a_0\} \cup \{\neg A\varphi \mid \neg A\varphi \in a_0\} \subseteq a.$$

Lemma 12 $a_0 \in \mathcal{A}$.

Our definition of relation $\prec_a$ in the canonical preference model is chosen to match item 4 of Definition 2.

Definition 5 For any agents $a, b, c \in \mathcal{A}$, let $b \prec_a c$ if there is a formula $L\varphi \in \Phi_0$ such that $L\varphi \in a$, $\varphi \notin b$, and $\varphi \in c$.

Note that the transitivity of the relation $\prec_a$ is non-obvious and, perhaps, surprising. We prove it as a part of the lemma below using the axioms of our logical system.

Lemma 13 *For each agent $a \in \mathcal{A}$, relation $b \prec_a c$ is a strict partial order on set $\mathcal{A}$.*

Proof We show the irreflexivity and the transitivity of relation $\prec_a$ separately:

Irreflexivity Suppose that there is an agent $b \in \mathcal{A}$ such that $b \prec_a b$. Thus, by Definition 5, there is a formula $L\varphi \in a$ such that $\varphi \notin b$ and $\varphi \in b$, which is a contradiction.

Transitivity Suppose that $b \prec_a c$ and $c \prec_a d$ for some agents $b, c, d \in \mathcal{A}$. Thus, by Definition 5, there are formulae $L\varphi \in \Phi_0$ and $L\psi \in \Phi_0$ such that

$$L\varphi \in a, \qquad \varphi \notin b, \qquad \varphi \in c. \qquad (9)$$

and

$$\mathsf{L}\psi \in a, \qquad \psi \notin c, \qquad \psi \in d. \tag{10}$$

By Definition 3, the assumptions $\mathsf{L}\varphi \in \Phi_0$ and $\mathsf{L}\psi \in \Phi_0$ imply that $\varphi, \psi \in \Phi_0$. Thus, by Eq. (8),

$$\mathsf{A}(\varphi \to \psi) \in \Phi_0^*, \qquad \mathsf{A}(\psi \to \varphi) \in \Phi_0^*, \tag{11}$$

$$\neg\mathsf{A}(\varphi \to \psi) \in \Phi_0^*, \qquad \neg\mathsf{A}(\psi \to \varphi) \in \Phi_0^*. \tag{12}$$

Note that by the Coherence of Preference axiom and propositional reasoning, the part $\mathsf{L}\varphi \in a$ of statement (9) and the part $\mathsf{L}\psi \in a$ of statement (10) imply

$$a \vdash \mathsf{A}(\varphi \to \psi) \lor \mathsf{A}(\psi \to \varphi).$$

Hence, either $\mathsf{A}(\varphi \to \psi) \in a$ or $\mathsf{A}(\psi \to \varphi) \in a$ by the maximality of set $a \subseteq \Phi_0^*$ and the statement (11). We consider these two cases separately:

Case I $\mathsf{A}(\varphi \to \psi) \in a$. Then, $\neg\mathsf{A}(\varphi \to \psi) \notin a$ because set a is consistent. Thus, $\neg\mathsf{A}(\varphi \to \psi) \notin a_0$ by Definition 4 and the assumption $a \in \mathcal{A}$ of the lemma. Hence, $\mathsf{A}(\varphi \to \psi) \in a_0$ by statements (11) and (12) and because a_0 is a maximal consistent subset of Φ_0^*. Then, $\mathsf{A}(\varphi \to \psi) \in c$ by Definition 4 and the assumption $c \in \mathcal{A}$. Then, $c \vdash \varphi \to \psi$ by the Reflexivity axiom and the Modus Ponens inference rule. Thus, $c \vdash \psi$ by part $\varphi \in c$ of statement (9). Note that the assumption $\mathsf{L}\psi \in \Phi_0$ implies that $\psi \in \Phi_0$ because by Definition 3 set Φ_0 is closed with respect to subformulae. Therefore, $\psi \in c$ because set c is a maximal consistent subset of Φ_0, which contradicts part $\psi \notin c$ of statement (10).

Case II $\mathsf{A}(\psi \to \varphi) \in a$. Then, $\neg\mathsf{A}(\psi \to \varphi) \notin a$ because set a is consistent. Thus, $\neg\mathsf{A}(\psi \to \varphi) \notin a_0$ by Definition 4 and the assumption $a \in \mathcal{A}$ of the lemma. Hence, $\mathsf{A}(\psi \to \varphi) \in a_0$ by statements (11) and (12) and because a_0 is a maximal consistent subset of Φ_0^*. Then, $\mathsf{A}(\psi \to \varphi) \in d$ by Definition 4 and the assumption $d \in \mathcal{A}$. Then, $d \vdash \psi \to \varphi$ by the Reflexivity axiom and the Modus Ponens inference rule. Thus, $d \vdash \varphi$ by part $\psi \in d$ of statement (10). Note that the assumption $\mathsf{L}\varphi \in \Phi_0$ implies that $\varphi \in \Phi_0$ because by Definition 3 set Φ_0 is closed with respect to subformulae. Hence, $\varphi \in d$ because set d is a maximal consistent subset of Φ_0. Therefore, $b \prec_a d$ by Definition 5 using parts $\mathsf{L}\varphi \in a$ and $\varphi \notin b$ of statement (9). $\qquad\square$

Definition 6 $\pi(p) = \{a \in \mathcal{A} \mid p \in a\}$ for each propositional variable $p \in \mathcal{P}$.

The canonical preference model $N(\Phi_0, a_0) = (\mathcal{A}, \{\prec_a\}_{a \in \mathcal{A}}, \pi)$ is now completely defined.

6.3 Auxiliary lemma

At the core of our proof of completeness is "truth" (or "induction") Lemma 15. In this subsection, we prove an auxiliary lemma used during the induction step of the proof of Lemma 15.

Lemma 14 *For any agent $a \in \mathcal{A}$ and any formula $\mathsf{L}\varphi \in \Phi_0$, if $\mathsf{L}\varphi \notin a$, then there are agents $b, c \in \mathcal{A}$ such that $\varphi \notin b$, $\varphi \in c$, and $b \not\preceq_a c$.*

Proof Outline. To prove the existence of agents b and c, we construct two consistent sets, $X_{k'}$ and Y and define agents c and b as maximal consistent extensions of sets $X_{k'}$ and Y, respectively.

Defining sets $X_{k'}$ and Y is not a trivial task. First, we introduce a total order $\sqsubseteq$ on a certain class of formulae Δ. Then, we use this total order to define a finite sequence of sets of formulae $X_0, X_1, \ldots, X_n$. Finally, we choose set $X_{k'}$ to be the last consistent set in this sequence. We use index k' of set $X_{k'}$ in the sequence $X_0, X_1, \ldots, X_n$ to define set Y.

Once sets $X_{k'}$ and Y are chosen and extended to maximal consistent sets c and b, we verify that the two later sets satisfy the conditions of the lemma.

Proof Consider the set of formulae

$$\Delta = \{\delta \in \Phi_0 \mid \mathsf{L}\delta \in a\}. \tag{13}$$

Next, we define a binary relation $\sqsubseteq$ on set Δ. For any formulae $\delta, \delta' \in \Delta$, let

$$\delta \sqsubseteq \delta' \text{ iff } \mathsf{A}(\delta \to \delta') \in a. \tag{14}$$

Claim Relation $\sqsubseteq$ is a total preorder on set Δ.
Proof of Claim We need to show that the relation is reflexive, transitive, and total.

Reflexivity Consider any formula $\delta \in \Delta$. It suffices to show that $\delta \sqsubseteq \delta$. Indeed, the formula $\delta \to \delta$ is a propositional tautology. Thus, by the Necessitation inference rule,

$$\vdash \mathsf{A}(\delta \to \delta). \tag{15}$$

At the same time, the assumption $\delta \in \Delta$ implies that $\delta \in \Phi_0$ by Eq. (13). Hence, $\mathsf{A}(\delta \to \delta) \in \Phi_0^*$ by Eq. (8). Then, $\mathsf{A}(\delta \to \delta) \in a$ by statement (15) and because a is a maximal consistent subset of Φ_0^*. Therefore, $\delta \sqsubseteq \delta$ by statement (14).

Transitivity Consider any formulae $\delta_1, \delta_2, \delta_3 \in \Delta$ such that $\delta_1 \sqsubseteq \delta_2$ and $\delta_2 \sqsubseteq \delta_3$. It suffices to show that $\delta_1 \sqsubseteq \delta_3$. Indeed, by statement (14), the assumptions $\delta_1 \sqsubseteq \delta_2$ and $\delta_2 \sqsubseteq \delta_3$ imply that

$$\mathsf{A}(\delta_1 \to \delta_2) \in a \quad \text{and} \quad \mathsf{A}(\delta_2 \to \delta_3) \in a. \tag{16}$$

Note that the formula

$$(\delta_1 \to \delta_2) \to ((\delta_2 \to \delta_3) \to (\delta_1 \to \delta_3))$$

is a propositional tautology. Thus, by the Necessitation inference rule,

$$\vdash \mathsf{A}((\delta_1 \to \delta_2) \to ((\delta_2 \to \delta_3) \to (\delta_1 \to \delta_3))).$$

Hence, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{A}(\delta_1 \to \delta_2) \to \mathsf{A}((\delta_2 \to \delta_3) \to (\delta_1 \to \delta_3)).$$

Then, by the Modus Ponens inference rule and the part $\mathsf{A}(\delta_1 \to \delta_2) \in a$ of statement (16),

$$a \vdash \mathsf{A}((\delta_2 \to \delta_3) \to (\delta_1 \to \delta_3)).$$

Thus, by the Distributivity axiom and the Modus Ponens inference rule,

$$a \vdash \mathsf{A}(\delta_2 \to \delta_3) \to \mathsf{A}(\delta_1 \to \delta_3).$$

Hence, by the Modus Ponens inference rule and part $\mathsf{A}(\delta_2 \to \delta_3) \in a$ of statement (16),

$$a \vdash \mathsf{A}(\delta_1 \to \delta_3). \tag{17}$$

The assumption $\delta_1, \delta_3 \in \Delta$ implies that $\delta_1, \delta_3 \in \Phi_0$ by Eq. (13). Hence, we have $\mathsf{A}(\delta_1 \to \delta_3) \in \Phi_0^*$ by Eq. (8). Then, $\mathsf{A}(\delta_1 \to \delta_3) \in a$ by statement (17) and because a is a maximal consistent subset of Φ_0^*. Therefore, $\delta_1 \sqsubseteq \delta_3$ by statement (14).

Totality Consider any formulae $\delta_1, \delta_2 \in \Delta$. It suffices to show that either $\delta_1 \sqsubseteq \delta_2$ or $\delta_2 \sqsubseteq \delta_1$. Indeed, by Eq. (13), the assumption $\delta_1, \delta_2 \in \Delta$ implies that $\mathsf{L}\delta_1 \in a$ and $\mathsf{L}\delta_2 \in a$. Thus, by the Coherence of Preferences axiom and propositional reasoning,

$$a \vdash \mathsf{A}(\delta_1 \to \delta_2) \vee \mathsf{A}(\delta_2 \to \delta_1). \tag{18}$$

At the same time, the assumption $\delta_1, \delta_2 \in \Delta$ implies $\delta_1, \delta_2 \in \Phi_0$ by Eq. (13). Hence, $\mathsf{A}(\delta_1 \to \delta_2) \in \Phi_0^*$ and $\mathsf{A}(\delta_2 \to \delta_1) \in \Phi_0^*$ by Eq. (8). Then, statement (18) implies that either $\mathsf{A}(\delta_1 \to \delta_2) \in a$ or $\mathsf{A}(\delta_2 \to \delta_1) \in a$, because set a is a maximal consistent subset of Φ_0^*. Therefore, either $\delta_1 \sqsubseteq \delta_2$ or $\delta_2 \sqsubseteq \delta_1$ by statement (14).

Note that set Δ is finite by Eq. (13) and because set Φ_0 is finite. Let

$$\delta_1 \sqsubseteq \delta_2 \sqsubseteq \cdots \sqsubseteq \delta_n \tag{19}$$

be any ordering of set Δ with respect to relation $\sqsubseteq$. Note that there might exist formulae $\delta, \delta' \in \Delta$ such that $\delta \sqsubseteq \delta'$ and $\delta' \sqsubseteq \delta$. Thus, ordering (19) might be not unique. We fix any such ordering.

For any integer k, such that $0 \le k \le n$, define a set of formulae

$$X_k = \{\varphi\} \cup \{\neg\delta_1, \ldots, \neg\delta_k\} \cup \{\mathsf{A}\psi \mid \mathsf{A}\psi \in a_0\} \cup \{\neg\mathsf{A}\chi \mid \neg\mathsf{A}\chi \in a_0\}. \tag{20}$$

Claim Set X_0 is consistent.
Proof of Claim Suppose the opposite. Thus, there are formulae

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_t \in a_0 \tag{21}$$

such that

$$A\psi_1, \ldots, A\psi_m, \neg A\chi_1, \ldots, \neg A\chi_t \vdash \neg\varphi.$$

Hence, by Lemma 6,

$$AA\psi_1, \ldots, AA\psi_m, A\neg A\chi_1, \ldots, A\neg A\chi_t \vdash A\neg\varphi.$$

Then, by Lemma 5 and the Modus Ponens inference rule applied m times,

$$A\psi_1, \ldots, A\psi_m, A\neg A\chi_1, \ldots, A\neg A\chi_t \vdash A\neg\varphi.$$

Thus, by the Euclidianity axiom and the Modus Ponens inference rule applied t times,

$$A\psi_1, \ldots, A\psi_m, \neg A\chi_1, \ldots, \neg A\chi_t \vdash A\neg\varphi.$$

Hence, by assumption (21),

$$a_0 \vdash A\neg\varphi. \tag{22}$$

The assumption $L\varphi \in \Phi_0$ of the lemma implies that $\varphi \in \Phi_0$ by Definition 3. Hence, $A\neg\varphi \in \Phi_0^*$ by Eq. (8). Then, $A\neg\varphi \in a_0$ by statement (22) and because set a_0 is a maximal consistent subset of Φ_0^*. Thus, $A\neg\varphi \in a$ by Definition 4. Hence, by the definition of modality A,

$$\neg\varphi \wedge L\neg\varphi \wedge L\neg\neg\varphi \in a. \tag{23}$$

Note that $\neg\neg\varphi \leftrightarrow \varphi$ is a propositional tautology. Thus, $\vdash A(\neg\neg\varphi \leftrightarrow \varphi)$ by the Necessitation inference rule. Hence, $\vdash L\neg\neg\varphi \to L\varphi$ by the Substitution axiom and the Modus Ponens inference rule. Hence, $a \vdash L\varphi$ by Eq. (23) and propositional reasoning. Therefore, $L\varphi \in a$ by the assumption $L\varphi \in \Phi_0 \subseteq \Phi_0^*$ of the lemma and because a is a maximal consistent subset of Φ_0^*, which contradicts the assumption $L\varphi \notin a$ of the lemma.

Set X_0 is consistent by the claim above. Let k' to be the maximal $k \leq n$ such that X_k is consistent. Consider the set of formulae

$$Y = \{\neg\varphi\} \cup \{\delta_{k'+1}, \ldots, \delta_n\} \cup \{A\psi \mid A\psi \in a_0\} \cup \{\neg A\chi \mid \neg A\chi \in a_0\}. \tag{24}$$

Claim Set Y is consistent.
Proof of Claim We consider the following two cases separately:
Case I $k' < n$. Then, set $X_{k'+1}$ is inconsistent by the choice of integer k'. Suppose that Y is also inconsistent. Thus, there are formulae

$$A\psi_1, \ldots, A\psi_m, \neg A\chi_1, \ldots, \neg A\chi_s \in a_0 \tag{25}$$
$$A\psi_1', \ldots, A\psi_\ell', \neg A\chi_1', \ldots, \neg A\chi_t' \in a_0 \tag{26}$$

such that

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_s, \neg\delta_{k'+1}, \neg\delta_{k'}, \neg\delta_{k'-1}, \ldots, \neg\delta_1 \vdash \neg\varphi,$$
$$\mathsf{A}\psi_1', \ldots, \mathsf{A}\psi_\ell', \neg\mathsf{A}\chi_1', \ldots, \neg\mathsf{A}\chi_t', \delta_{k'+1}, \delta_{k'+2}, \delta_{k'+3}, \ldots, \delta_n \vdash \varphi.$$

Hence, by the Modus Ponens inference rule applied k' times to the first statement and $(n - k - 1)$ times to the second one,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_s,$$
$$\neg\delta_{k'+1}, \neg\delta_{k'+1} \to \neg\delta_{k'}, \neg\delta_{k'} \to \neg\delta_{k'-1}, \ldots, \neg\delta_2 \to \neg\delta_1 \vdash \neg\varphi,$$

$$\mathsf{A}\psi_1', \ldots, \mathsf{A}\psi_\ell', \neg\mathsf{A}\chi_1', \ldots, \neg\mathsf{A}\chi_t',$$
$$\delta_{k'+1}, \delta_{k'+1} \to \delta_{k'+2}, \delta_{k'+2} \to \delta_{k'+3}, \ldots, \delta_{n-1} \to \delta_n \vdash \varphi.$$

Then, by Lemma 4 applied to both statements,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_s,$$
$$\neg\delta_{k'+1} \to \neg\delta_{k'}, \neg\delta_{k'} \to \neg\delta_{k'-1}, \ldots, \neg\delta_2 \to \neg\delta_1 \vdash \neg\delta_{k'+1} \to \neg\varphi,$$

$$\mathsf{A}\psi_1', \ldots, \mathsf{A}\psi_\ell', \neg\mathsf{A}\chi_1', \ldots, \neg\mathsf{A}\chi_t',$$
$$\delta_{k'+1} \to \delta_{k'+2}, \delta_{k'+2} \to \delta_{k'+3}, \ldots, \delta_{n-1} \to \delta_n \vdash \delta_{k'+1} \to \varphi.$$

Thus, by the law of contraposition applied $(k' + 1)$ times to the first statement,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_s,$$
$$\delta_{k'} \to \delta_{k'+1}, \delta_{k'-1} \to \delta_{k'}, \ldots, \delta_1 \to \delta_2 \vdash \varphi \to \delta_{k'+1},$$

$$\mathsf{A}\psi_1', \ldots, \mathsf{A}\psi_\ell', \neg\mathsf{A}\chi_1', \ldots, \neg\mathsf{A}\chi_t',$$
$$\delta_{k'+1} \to \delta_{k'+2}, \delta_{k'+2} \to \delta_{k'+3}, \ldots, \delta_{n-1} \to \delta_n \vdash \delta_{k'+1} \to \varphi.$$

Hence, by Lemma 6 applied to both statements,

$$\mathsf{A}\mathsf{A}\psi_1, \ldots, \mathsf{A}\mathsf{A}\psi_m, \mathsf{A}\neg\mathsf{A}\chi_1, \ldots, \mathsf{A}\neg\mathsf{A}\chi_s,$$
$$\mathsf{A}(\delta_{k'} \to \delta_{k'+1}), \mathsf{A}(\delta_{k'-1} \to \delta_{k'}), \ldots, \mathsf{A}(\delta_1 \to \delta_2) \vdash \mathsf{A}(\varphi \to \delta_{k'+1}),$$

$$\mathsf{A}\mathsf{A}\psi_1', \ldots, \mathsf{A}\mathsf{A}\psi_\ell', \mathsf{A}\neg\mathsf{A}\chi_1', \ldots, \mathsf{A}\neg\mathsf{A}\chi_t',$$
$$\mathsf{A}(\delta_{k'+1} \to \delta_{k'+2}), \mathsf{A}(\delta_{k'+2} \to \delta_{k'+3}), \ldots, \mathsf{A}(\delta_{n-1} \to \delta_n) \vdash \mathsf{A}(\delta_{k'+1} \to \varphi).$$

Then, by Lemma 5 and the Modus Ponens inference rule applied m times to the first statement and ℓ times to the second statement,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \mathsf{A}\neg\mathsf{A}\chi_1, \ldots, \mathsf{A}\neg\mathsf{A}\chi_s,$$

$$A(\delta_{k'} \to \delta_{k'+1}), A(\delta_{k'-1} \to \delta_{k'}), \ldots, A(\delta_1 \to \delta_2) \vdash A(\varphi \to \delta_{k'+1}),$$

$$A\psi_1', \ldots, A\psi_\ell', A\neg A\chi_1', \ldots, A\neg A\chi_t',$$
$$A(\delta_{k'+1} \to \delta_{k'+2}), A(\delta_{k'+2} \to \delta_{k'+3}), \ldots, A(\delta_{n-1} \to \delta_n) \vdash A(\delta_{k'+1} \to \varphi).$$

Thus, by the Euclidianity axiom and the Modus Ponens inference rule applied s times to the first statement and t times to the second statement,

$$A\psi_1, \ldots, A\psi_m, \neg A\chi_1, \ldots, \neg A\chi_s,$$
$$A(\delta_{k'} \to \delta_{k'+1}), A(\delta_{k'-1} \to \delta_{k'}), \ldots, A(\delta_1 \to \delta_2) \vdash A(\varphi \to \delta_{k'+1}),$$

$$A\psi_1', \ldots, A\psi_\ell', \neg A\chi_1', \ldots, \neg A\chi_t',$$
$$A(\delta_{k'+1} \to \delta_{k'+2}), A(\delta_{k'+2} \to \delta_{k'+3}), \ldots, A(\delta_{n-1} \to \delta_n) \vdash A(\delta_{k'+1} \to \varphi).$$

Observe that $A\psi_1, \ldots, A\psi_m, \neg A\chi_1, \ldots, \neg A\chi_s \in a$ by Definition 4 and the assumptions $a \in \mathcal{A}$ and (25). Similarly, $A\psi_1', \ldots, A\psi_\ell', \neg A\chi_1', \ldots, \neg A\chi_t' \in a$ by the assumption (26). Hence,

$$a, A(\delta_{k'} \to \delta_{k'+1}), A(\delta_{k'-1} \to \delta_{k'}), \ldots, A(\delta_1 \to \delta_2) \vdash A(\varphi \to \delta_{k'+1}),$$
$$a, A(\delta_{k'+1} \to \delta_{k'+2}), A(\delta_{k'+2} \to \delta_{k'+3}), \ldots, A(\delta_{n-1} \to \delta_n) \vdash A(\delta_{k'+1} \to \varphi).$$

Recall that $\delta_1 \sqsubseteq \delta_2 \sqsubseteq \cdots \sqsubseteq \delta_n$ by assumption (19). Then,

$$A(\delta_1 \to \delta_2), \ldots, A(\delta_{n-1} \to \delta_n) \in a$$

by statement (14). Thus,

$$a \vdash A(\varphi \to \delta_{k'+1}), \tag{27}$$
$$a \vdash A(\delta_{k'+1} \to \varphi). \tag{28}$$

Note that formula

$$(\varphi \to \delta_{k'+1}) \to ((\delta_{k'+1} \to \varphi) \to (\delta_{k'+1} \leftrightarrow \varphi))$$

is a propositional tautology. Hence,

$$\varphi \to \delta_{k'+1}, \delta_{k'+1} \to \varphi \vdash \delta_{k'+1} \leftrightarrow \varphi$$

by the Modus Ponens rule applied twice. Then,

$$A(\varphi \to \delta_{k'+1}), A(\delta_{k'+1} \to \varphi) \vdash A(\delta_{k'+1} \leftrightarrow \varphi)$$

by Lemma 6. Thus, $a \vdash \mathsf{A}(\delta_{k'+1} \leftrightarrow \varphi)$ by statements (27) and (28). Hence, by the Substitution axiom and the Modus Ponens inference rule,

$$a \vdash \mathsf{L}\delta_{k'+1} \to \mathsf{L}\varphi.$$

Equation (13) implies that $\mathsf{L}\delta_{k'+1} \in a$ because $\delta_{k'+1} \in \Delta$. Then,

$$a \vdash \mathsf{L}\varphi.$$

Recall that $\mathsf{L}\varphi \in \Phi_0 \subseteq \Phi_0^*$ by the assumption of the lemma. Thus, $\mathsf{L}\varphi \in a$ because a is a maximal consistent subset of Φ_0^*, which contradicts the assumption $\mathsf{L}\varphi \notin a$ of the lemma.

Case II $k' = n$. Suppose that set Y is not consistent. Thus, there are formulae

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_t \in a_0 \tag{29}$$

such that

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_t \vdash \varphi.$$

Hence, by Lemma 6,

$$\mathsf{AA}\psi_1, \ldots, \mathsf{AA}\psi_m, \mathsf{A}\neg\mathsf{A}\chi_1, \ldots, \mathsf{A}\neg\mathsf{A}\chi_t \vdash \mathsf{A}\varphi.$$

Then, by Lemma 5 and the Modus Ponens inference rule applied m times,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \mathsf{A}\neg\mathsf{A}\chi_1, \ldots, \mathsf{A}\neg\mathsf{A}\chi_t \vdash \mathsf{A}\varphi.$$

Thus, by the Euclidianity axiom and the Modus Ponens rule applied t times,

$$\mathsf{A}\psi_1, \ldots, \mathsf{A}\psi_m, \neg\mathsf{A}\chi_1, \ldots, \neg\mathsf{A}\chi_t \vdash \mathsf{A}\varphi.$$

Hence, by assumption (29),

$$a_0 \vdash \mathsf{A}\varphi.$$

Recall that $\mathsf{L}\varphi \in \Phi_0$ by the assumption of the lemma. Hence, $\varphi \in \Phi_0$ by Definition 3. Then, $\mathsf{A}\varphi \in \Phi_0^*$ by Eq. (8). Thus, $\mathsf{A}\varphi \in a_0$ because a_0 is a maximal consistent subset of Φ_0^*. Hence, $\mathsf{A}\varphi \in a$ by Definition 4 and the assumption $a \in \mathcal{A}$. Then, $\varphi \wedge \mathsf{L}\varphi \wedge \mathsf{L}\neg\varphi \in a$ by the definition of modality A. Hence, by propositional reasoning,

$$a \vdash \mathsf{L}\varphi.$$

Recall that $\mathsf{L}\varphi \in \Phi_0 \subseteq \Phi_0^*$ by the assumption of the lemma. Thus, $\mathsf{L}\varphi \in a$ because a is a maximal consistent subset of Φ_0^*, which contradicts the assumption $\mathsf{L}\varphi \notin a$ of the lemma.

Claim $X_{k'}, Y \subseteq \Phi_0^*$.

Proof of Claim Consider any formula $\zeta \in X_{k'} \cup Y$. It suffices to show that $\zeta \in \Phi_0^*$. By Eqs. (20) and (24), it is enough to consider the following three cases:

Case I $\zeta \equiv \varphi$ or $\zeta \equiv \neg\varphi$. Assumption $\mathsf{L}\varphi \in \Phi_0$ of the lemma implies that $\neg\varphi \in \Phi_0$ by item 3 of Definition 3. Then, $\varphi \in \Phi_0$ because set Φ_0 is closed with respect to subformulae. Therefore, $\varphi, \neg\varphi \in \Phi_0 \subseteq \Phi_0^*$.

Case II $\zeta \equiv \delta$ or $\zeta \equiv \neg\delta$ for some formula $\delta \in \Delta$. Then, $\mathsf{L}\delta \in a$ by Eq. (13). Thus, $\neg\delta \in \Phi_0$ by item 3 of Definition 3. Hence, $\delta \in \Phi_0$ because set Φ_0 is closed with respect to subformulae. Therefore, $\delta, \neg\delta \in \Phi_0 \subseteq \Phi_0^*$.

Case III $\zeta \in a_0$. Then, $\zeta \in \Phi_0^*$ because set a_0 is a subset of Φ_0^*.

Let sets b and c be any maximal consistent subsets of Φ_0^* that contain sets Y and $X_{k'}$, respectively. Such subsets exist by the above claim and Lemma 3. Note that $b, c \in \mathcal{A}$ by Definition 4 and Eqs. (24) and (20). Also, $\neg\varphi \in Y \subseteq b$ by Eq. (24). Then, $\varphi \notin b$ because set b is consistent. Additionally, $\varphi \in X_{k'} \subseteq c$ by Eq. (20).

Claim $b \not\prec_a c$.

Proof of Claim Suppose that $b \prec_a c$. Thus, by Definition 5, there is a formula $\mathsf{L}\delta \in \Phi_0$ such that $\mathsf{L}\delta \in a$,

$$\delta \notin b, \tag{30}$$

and

$$\delta \in c. \tag{31}$$

Hence, $\delta \in \Delta$ by Eq. (13) and because set Φ_0 is closed with respect to subformulae. Recall that chain (19) is the ordering of all formulae in Δ. Let $\delta = \delta_i$ for some $i \leq n$. We consider the following two cases separately:

Case I $i \leq k'$. Then, $\neg\delta = \neg\delta_i \in X_{k'} \subseteq c$ by Eq. (20) and the choice of set c. Hence, $\delta \notin c$ because set c is consistent, which contradicts statement (31).

Case II $i > k'$. Then, $\delta = \delta_i \in Y \subseteq b$ by Eq. (24) and the choice of set b, which contradicts statement (30). $\qquad\square$

This concludes the proof of the lemma.

6.4 Final steps

Lemma 15 $a \Vdash \varphi$ *iff* $\varphi \in a$ *for each agent* $a \in \mathcal{A}$ *and each formula* $\varphi \in \Phi_0$.

Proof We prove the lemma by induction on the structural complexity of formula φ. If formula φ is a propositional variable, then the required follows from item 1 of Definition 2 and Definition 6. If formula φ is a negation or an implication, then the required follows from items 2 and 3 of Definition 2, the induction hypothesis, and the maximality and the consistency of the set a in the standard way.

Finally, let formula φ have the form $\mathsf{L}\psi$.

$(\Rightarrow)$: Assume that $\mathsf{L}\psi \notin a$. Then, by Lemma 14, there are agents $b, c \in \mathcal{A}$ such that $\psi \notin b$, $\psi \in c$, and $b \not\prec_a c$. Thus, $b \not\Vdash \psi$ and $c \Vdash \psi$ by the induction hypothesis. Therefore, $a \not\Vdash \mathsf{L}\psi$ by item 4 of Definition 2.

($\Longleftarrow$) : Assume that $\mathsf{L}\psi \in a$. Consider any agents b, $c \in \mathcal{A}$ such that $b \nVdash \psi$ and $c \Vdash \psi$. By item 4 of Definition 2 it suffices to show that $b \prec_a c$. Indeed, by the induction hypothesis, assumptions $b \nVdash \psi$ and $c \Vdash \psi$ imply that $\psi \notin b$ and $\psi \in c$. Then, $b \prec_a c$ by Definition 5 and the assumption $\mathsf{L}\psi \in a$ of the case. $\square$

We are now ready to prove the completeness of our logical system. Recall that we used the filtration technique to guarantee that the set of agents is finite. As a result, we can prove the completeness, but not strong completeness of our system.

Theorem 2 (completeness) *If $a \Vdash \varphi$ for each agent a of each preference model, then $\vdash \varphi$.*

Proof Suppose that $\nvdash \varphi$. Consider the singleton set $\{\neg\varphi\}$. By Lemma 11, it can be extended to a finite saturated set Φ_0. Let Φ_0^* be the extension of Φ_0 defined by Eq. (8). Assumption $\nvdash \varphi$ implies that set $\{\neg\varphi\}$ is consistent. Then by Lemma 3, there is a maximal consistent subset $a_0 \subseteq \Phi_0^*$ such that $\{\neg\varphi\} \subseteq a_0$. Consider the canonical preference model $N(\Phi_0, a_0)$. By Lemma 12, set a_0 is an agent in the network $N(\Phi_0, a_0)$. Note that, by the choice of set a_0, we have $\neg\varphi \in \{\neg\varphi\} \subseteq a_0$. Thus, $\varphi \notin a_0$ because set a_0 is consistent. Therefore, $a_0 \nVdash \varphi$ by Lemma 15. $\square$

We conclude this section with an observation that our logical system is decidable, which follows from the completeness theorem.

Theorem 3 *Set $\{\varphi \in \Phi \mid \vdash \varphi\}$ is decidable.*

Proof The set $\{\varphi \in \Phi \mid \vdash \varphi\}$ is recursively enumerable because it is axiomatisable. Also, the set $\{\varphi \in \Phi \mid \nvdash \varphi\}$ is recursively enumerable because our logical system is sound and complete with respect to the preference models with finite sets of agents. Therefore, the set $\{\varphi \in \Phi \mid \vdash \varphi\}$ is decidable. $\square$

7 The logic of conditional likes

In the previous sections, we considered such *unconditional* likes statements as "*a* likes logicians" and "*a* likes philosophers". In this section, following Lang et al. (2002), we consider *conditional* likes statements such as "among all philosophers, *a* likes logicians". We write the last statement as

$$a \Vdash \mathsf{L}(\text{"is a logician"} \mid \text{"is a philosopher"}).$$

Below, we formally introduce the syntax and the semantics for a language with conditional likes modality $\mathsf{L}(\cdot \mid \cdot)$ and discuss possible axioms for conditional likes. We also show that unconditional likes modality L is definable through conditional likes modality but not the other way around.

7.1 Syntax and semantics

The language Φ^+ that we consider in this section is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{L}(\varphi \mid \varphi).$$

We read $\mathsf{L}(\varphi \mid \psi)$ as "likes those who have property φ, among those who satisfy condition ψ". We again assume that disjunction $\vee$, conjunction $\wedge$, and biconditional $\leftrightarrow$ are defined in our language in the standard way. We also assume the same about constant $\top$.

The semantics of formulae in language Φ^+ could be defined by replacing item 4 of Definition 2 with

$$a \Vdash \mathsf{L}(\varphi \mid \psi), \; when \, \forall b, c \; (if \; b \Vdash \neg\varphi \wedge \psi \; and \; c \Vdash \varphi \wedge \psi, \; then \; b \prec_a c). \quad (32)$$

Note that the conditions $b \Vdash \neg\varphi \wedge \psi$ and $c \Vdash \varphi \wedge \psi$ imply that $b \neq c$. Thus, just like it was in the case of unconditional likes, a replacement of the strict relation $\prec_a$ in the above statement with the no-strict relation $\preceq_a$ results in an equivalent definition of the conditional likes modality. See our discussion after Definition 2. Observe also that formulae $\mathsf{L}(\varphi \mid \top)$ is equivalent under the above semantics to the formula $\mathsf{L}\varphi$. Thus, the logical system that we consider in this section could be considered to be an extension of the logical system for unconditional likes modality from the previous section. As before, "for all agents" modality $\mathsf{A}\varphi$ can be defined as $\varphi \wedge \mathsf{L}(\varphi \mid \top) \wedge \mathsf{L}(\neg\varphi \mid \top)$.

7.2 Axioms

Below are some of the universally true properties that can be expressed in the language Φ^+.

1. $\mathsf{L}(\varphi_1 \mid \psi_1) \wedge \mathsf{L}(\varphi_2 \mid \psi_2) \rightarrow \mathsf{L}(\varphi_1 \wedge \varphi_2 \mid \psi_1 \wedge \psi_2)$,
2. $\mathsf{L}(\varphi_1 \mid \psi_1) \wedge \mathsf{L}(\varphi_2 \mid \psi_2) \rightarrow \mathsf{L}(\varphi_1 \vee \varphi_2 \mid \psi_1 \wedge \psi_2)$,
3. $\mathsf{L}(\varphi_1 \wedge \varphi_2 \mid \psi) \rightarrow \mathsf{L}(\varphi_1 \mid \varphi_2 \wedge \psi)$,
4. $\mathsf{L}(\varphi_1 \vee \varphi_2 \mid \psi) \rightarrow \mathsf{L}(\varphi_1 \mid \neg\varphi_2 \wedge \psi)$.

Note, however, that unlike unconditional likes modality, which has a relatively simple axiomatisation, conditional likes modality has much less trivial and harder-to-express properties. An example of them is

$$\bigwedge_{0 \leq i < n} \mathsf{L}(\varphi_i \mid \psi_i) \rightarrow \bigvee_{0 \leq i < n} \mathsf{A}\neg(\varphi_i \wedge \psi_i \wedge \neg\varphi_{i'} \wedge \psi_{i'}),$$

where, for each i, integer $i' < n$ is such that $i' \equiv i + 1 \pmod{n}$. We believe that this property does not follow from properties 1–4 listed in the beginning of this subsection and that it cannot be reduce to a combination of some simple principles.

We prove the soundness of this principle in the theorem below. A complete axiomatisation of modality $\mathsf{L}(\varphi \mid \psi)$ remains an open question.

Theorem 4 *For an arbitrary preference model $(\mathcal{A}, \{\prec_a\}_{a \in \mathcal{A}}, \pi)$, any agent $a \in \mathcal{A}$, any $n \geq 1$, and any formulae $\varphi_0, \ldots, \varphi_{n-1}, \psi_0, \ldots, \psi_{n-1}$, if $a \Vdash \mathsf{L}(\varphi_i \mid \psi_i)$ for each $i < n$, then there is $i < n$ such that*

$$a \Vdash \mathsf{A}\neg(\varphi_i \wedge \psi_i \wedge \neg\varphi_{i'} \wedge \psi_{i'}),$$

where $i' < n$ is such that $i' \equiv i + 1 \pmod{n}$.

Proof Towards contradiction, suppose that

$$a \Vdash \mathsf{L}(\varphi_i \mid \psi_i), \qquad \forall i < n \tag{33}$$

and

$$a \nVdash \mathsf{A}\neg(\varphi_i \wedge \psi_i \wedge \neg\varphi_{i'} \wedge \psi_{i'}),$$

for each $i < n$, where $i' < n$ is such that $i' \equiv i+1 \pmod{n}$. Thus, by Lemma 1, for each $i < n$ there is an agent $b_i \in \mathcal{A}$ such that

$$b_i \nVdash \neg(\varphi_i \wedge \psi_i \wedge \neg\varphi_{i'} \wedge \psi_{i'}),$$

where $i' < n$ is such that $i' \equiv i+1 \pmod{n}$. Hence, by item 2 of Definition 2,

$$b_i \Vdash \varphi_i \wedge \psi_i \wedge \neg\varphi_{i'} \wedge \psi_{i'},$$

where $i' < n$ is such that $i' \equiv i+1 \pmod{n}$. Then,

$$\begin{array}{ll} b_0 \Vdash \varphi_0 \wedge \psi_0, & b_0 \Vdash \neg\varphi_1 \wedge \psi_1, \\ b_1 \Vdash \varphi_1 \wedge \psi_1, & b_1 \Vdash \neg\varphi_2 \wedge \psi_2, \\ b_2 \Vdash \varphi_2 \wedge \psi_2, & b_2 \Vdash \neg\varphi_3 \wedge \psi_3, \\ \quad \cdots & \quad \cdots \\ b_{n-2} \Vdash \varphi_{n-2} \wedge \psi_{n-2}, & b_{n-2} \Vdash \neg\varphi_{n-1} \wedge \psi_{n-1}, \\ b_{n-1} \Vdash \varphi_{n-1} \wedge \psi_{n-1}, & b_{n-1} \Vdash \neg\varphi_0 \wedge \psi_0. \end{array}$$

These statements can be rearranged as,

$$\begin{array}{ll} b_0 \Vdash \neg\varphi_1 \wedge \psi_1, & b_1 \Vdash \varphi_1 \wedge \psi_1, \\ b_1 \Vdash \neg\varphi_2 \wedge \psi_2, & b_2 \Vdash \varphi_2 \wedge \psi_2, \\ b_2 \Vdash \neg\varphi_3 \wedge \psi_3, & b_3 \Vdash \varphi_3 \wedge \psi_3, \\ \quad \cdots & \quad \cdots \\ b_{n-2} \Vdash \neg\varphi_{n-1} \wedge \psi_{n-1}, & b_{n-1} \Vdash \varphi_{n-1} \wedge \psi_{n-1}, \\ b_{n-1} \Vdash \neg\varphi_0 \wedge \psi_0 & b_0 \Vdash \varphi_0 \wedge \psi_0. \end{array}$$

Thus, by assumption (33) and statement (32),

$$b_0 \prec_a b_1 \prec_a b_2 \prec_a \cdots \prec_a b_{n-1} \prec_a b_0.$$

Therefore, $b_0 \prec_a b_0$ because relation $\prec_a$ is transitive by Definition 1, which contradicts $\prec_a$ being a strict partial order by the same Definition 1. □

7.3 Undefinability

In this section, we show that conditional likes modality $L(\cdot \mid \cdot)$ is not definable in our original language Φ. To achieve this, we consider the preference model depicted in Fig. 2. It has four agents: a, b, c, and d. The only preference in the model is that agent c prefers agent a over agent d. Without loss of generality, we assume that the language has only two propositional variables: p and q. Finally, let $\pi(p) = \{a\}$ and $\pi(q) = \{a, d\}$. In other words, propositional variable p is true only about agent a and propositional variable q is true only about agents a and d, see Fig. 2. We will show that agents b and c are indistinguishable in language Φ and that they are distinguishable in a language containing conditional likes modality $L(\cdot \mid \cdot)$.

We start the proof with two auxiliary lemmas. The first of them is true because relation $\prec_b$ is empty.

Lemma 16 *If X and Y are any two sets of agents in the preference model depicted in Fig. 2 such that $X \cup Y = \{a, b, c, d\}$ and $x \prec_b y$ for all $x \in X$ and all $y \in Y$, then either set X or set Y is empty.*

The second auxiliary lemma is true because agents b and c are not in relation $\prec_c$ with any other agents.

Lemma 17 *If X and Y are any two sets of agents in the preference model depicted in Fig. 2 such that $X \cup Y = \{a, b, c, d\}$ and $x \prec_c y$ for all $x \in X$ and all $y \in Y$, then either set X or set Y is empty.*

Proof If $b \in X$, then set Y must be empty because $b \not\prec_c y$ for any $y \in \{a, b, c, d\}$, see Fig. 2. Similarly, if $b \in Y$, then set X must be empty. □

Lemma 18 $b \Vdash \varphi$ iff $c \Vdash \varphi$ for any formula $\varphi \in \Phi$.

Proof We prove the statement by induction on the structural complexity of formula φ. If formula φ is propositional variable p or q, then the statement holds by item 1 of Definition 2 and because $b, c \notin \pi(p)$ and $b, c \notin \pi(q)$. If formula φ is a negation or an implication, then the statement of the lemma follows from the induction hypothesis and either item 2 or item 3 of Definition 2 in the standard way.
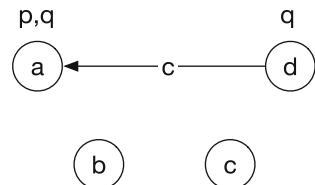
Suppose that formula φ has the form $L\psi$.

($\Longleftarrow$) : Suppose that $c \Vdash L\psi$. Consider sets

$$X = \{x \in \mathcal{A} \mid x \not\Vdash \psi\},$$
$$Y = \{y \in \mathcal{A} \mid y \Vdash \psi\}.$$

Fig. 2 A preference model

The assumption $c \Vdash \mathsf{L}\psi$ implies that $x \prec_c y$ for each agent $x \in X$ and each agent $y \in Y$ by item 4 of Definition 2. Then, either $X = \varnothing$ or $Y = \varnothing$ by Lemma 17. Hence, $b \Vdash \mathsf{L}\psi$ again by item 4 of Definition 2.

$(\Rightarrow)$: The proof is similar, but it uses Lemma 16 instead of Lemma 17. $\qquad\square$

The next lemma follows from statement (32) and the specification of the model in Fig. 2.

Lemma 19 $b \nVdash \mathsf{L}(p \mid q)$ and $c \Vdash \mathsf{L}(p \mid q)$.

The next theorem follows from the last two lemmas.

Theorem 5 *Modality* $\mathsf{L}(\cdot \mid \cdot)$ *is not definable in language* Φ.

8 Conclusion

We have proposed an egocentric approach to reasoning about higher-order preferences. Unlike the previous works on preference and desire logics, the egocentric approach allows us to express higher-order preferences using nested statements of the form "agent a likes those who like philosophers". We have considered unconditional and conditional modalities capturing such preferences. We have observed that unconditional modality is definable through conditional, but not the other way around. Our main technical result is a sound, complete, and decidable logical system describing the properties of the unconditional likes modality.

An interesting direction for future research is to introduce the distinction between *de re* and *de dicto* likes. An agent might like a group of people because they are philosophers or the group liked by the agent just happens to consist of philosophers.

Appendix

To keep this article self-contained, in this appendix, we prove the standard lemmas mentioned in Sect. 4.

Lemma 4 *If* $X, \varphi \vdash \psi$, *then* $X \vdash \varphi \to \psi$.

Proof Suppose that sequence $\psi_1, \ldots, \psi_n$ is a proof from set $X \cup \{\varphi\}$ and the theorems of our logical system that uses the Modus Ponens inference rule only. In other words, for each $k \le n$, either

1. $\vdash \psi_k$, or
2. $\psi_k \in X$, or
3. ψ_k is equal to φ, or
4. there are $i, j < k$ such that formula ψ_j is equal to $\psi_i \to \psi_k$.

It suffices to show that $X \vdash \varphi \to \psi_k$ for each $k \le n$. We prove this by induction on k through considering the four cases above separately.

Case I: $\vdash \psi_k$. Note that $\psi_k \to (\varphi \to \psi_k)$ is a propositional tautology, and thus, is an axiom of our logical system. Hence, $\vdash \varphi \to \psi_k$ by the Modus Ponens inference rule. Therefore, $X \vdash \varphi \to \psi_k$.

Case II: $\psi_k \in X$. Then, similar to the previous case, $X \vdash \varphi \to \psi_k$.

Case III: formula ψ_k is equal to φ. Thus, $\varphi \to \psi_k$ is a propositional tautology. Therefore, $X \vdash \varphi \to \psi_k$.

Case IV: formula ψ_j is equal to $\psi_i \to \psi_k$ for some i, $j < k$. Thus, by the induction hypothesis, $X \vdash \varphi \to \psi_i$ and $X \vdash \varphi \to (\psi_i \to \psi_k)$. Note that formula

$$(\varphi \to \psi_i) \to ((\varphi \to (\psi_i \to \psi_k)) \to (\varphi \to \psi_k))$$

is a propositional tautology. Therefore, $X \vdash \varphi \to \psi_k$ by applying the Modus Ponens inference rule twice. $\square$

Lemma 5 $\vdash \mathsf{A}\varphi \to \mathsf{A}\mathsf{A}\varphi$.

Proof Note that formula $\mathsf{A}\neg\mathsf{A}\varphi \to \neg\mathsf{A}\varphi$ is an instance of the Reflexivity axiom. Thus, by contraposition, $\vdash \mathsf{A}\varphi \to \neg\mathsf{A}\neg\mathsf{A}\varphi$. Then, taking into account that $\vdash \neg\mathsf{A}\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\neg\mathsf{A}\varphi$ by the Euclidianity axiom, one can conclude that

$$\vdash \mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\neg\mathsf{A}\varphi. \tag{34}$$

At the same time, $\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi$ is an instance of the Euclidianity axiom. Thus, $\vdash \neg\mathsf{A}\neg\mathsf{A}\varphi \to \mathsf{A}\varphi$ by contraposition. Hence, $\vdash \mathsf{A}(\neg\mathsf{A}\neg\mathsf{A}\varphi \to \mathsf{A}\varphi)$ by the Necessitation inference rule. Thus,

$$\vdash \mathsf{A}\neg\mathsf{A}\neg\mathsf{A}\varphi \to \mathsf{A}\mathsf{A}\varphi$$

by the Distributivity axiom and the Modus Ponens inference rule. The last statement, together with statement (34), implies the statement of the lemma by the laws of propositional reasoning. $\square$

Lemma 6 *If* $\varphi_1, \ldots, \varphi_n \vdash \psi$, *then* $\mathsf{A}\varphi_1, \ldots, \mathsf{A}\varphi_n \vdash \mathsf{A}\psi$.

Proof By Lemma 4 applied n times, the assumption $\varphi_1, \ldots, \varphi_n \vdash \psi$ implies that

$$\vdash \varphi_1 \to (\varphi_2 \to \ldots (\varphi_n \to \psi) \ldots).$$

Thus,

$$\vdash \mathsf{A}(\varphi_1 \to (\varphi_2 \to \ldots (\varphi_n \to \psi) \ldots))$$

by the Necessitation inference rule. Hence, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{A}\varphi_1 \to \mathsf{A}(\varphi_2 \to \ldots (\varphi_n \to \psi) \ldots).$$

Then, $\mathsf{A}\varphi_1 \vdash \mathsf{A}(\varphi_2 \to \ldots (\varphi_n \to \psi) \ldots)$, again by the Modus Ponens inference rule. Therefore, $\mathsf{A}\varphi_1, \ldots, \mathsf{A}\varphi_n \vdash \mathsf{A}\psi$ by applying the previous steps $(n-1)$ more times. $\square$

Declarations

References

Åqvist, L. (1962). A binary primitive in deontic logic. *Logique et Analyse, 5*(19), 90–97.

Baltag, A., & Smets, S. (2009). Talking your way into agreement: Belief merge by persuasive communication. In: M. Baldoni, C. Baroglio, J. Bentahar, G. Boella, M. Cossentino, M. Dastani, B. Dunin-Keplicz, G. Fortino, M. Gleizes, J. Leite, V. Mascardi, J.A. Padget, J. Pavón, A. Polleres, A.E.F. Seghrouchni, P. Torroni, R. Verbrugge (eds.) Proceedings of the Second Multi-Agent Logics, Languages, and Organisations Federated Workshops, Turin, Italy, September 7–10, 2009, *CEUR Workshop Proceedings*, vol. 494. CEUR-WS.org

Chisholm, R. M., & Sosa, E. (1966). On the logic of "intrinsically better". *American Philosophical Quarterly, 3*(3), 244–249.

Christoff, Z., Gratzl, N., & Roy, O. (2021). Priority merge and intersection modalities. *The Review of Symbolic Logic*. https://doi.org/10.1017/S1755020321000058

Christoff, Z., & Hansen, J. U. (2015). A logic for diffusion in social networks. *Journal of Applied Logic, 13*(1), 48–77.

Christoff, Z., Hansen, J. U., & Proietti, C. (2016). Reflecting on social influence in networks. *Journal of Logic, Language and Information, 25*(3), 299–333.

Goranko, V., & van Drimmelen, G. (2006). Complete axiomatization and decidability of alternating-time temporal logic. *Theoretical Computer Science, 353*(1), 93–117. https://doi.org/10.1016/j.tcs.2005.07.043

Grove, A. J. (1995). Naming and identity in epistemic logic part II: A first-order logic for naming. *Artificial Intelligence, 74*(2), 311–350.

Grove, A.J., & Halpern, J.Y. (1991) Naming and identity in a multi-agent epistemic logic. In: J.F. Allen, R. Fikes, E. Sandewall (eds.) Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Cambridge, MA, USA, April 22–25, 1991, pp. 301–312. Morgan Kaufmann.

Grove, A. J., & Halpern, J. Y. (1993). Naming and identity in epistemic logics part I: The propositional case. *Journal of Logic and Computation, 3*(4), 345–378.

Halldén, S. (1957). *On the logic of "better"*. Almqvist & Wiksells.

Jiang, J., & Naumov, P. (2002). The egocentric logic of preferences. In: the 31st International Joint Conference on Artificial Intelligence (IJCAI-22), (pp. 2676–2682).

Kawasaki, G. (2004). *The art of the start: The time-tested, battle-hardened guide for anyone starting anything*. Penguin.

Lang, J., van der Torre, L., & Weydert, E. (2002). Utilitarian desires. *Autonomous Agents and Multi-Agent Systems, 5*(3), 329–363.

Liu, F. (2011). *Reasoning about preference dynamics* (Vol. 354). Springer.

Lorini, E. (2021). A qualitative theory of cognitive attitudes and their change. *Theory and Practice of Logic Programming, 21*(4), 428–458.

Mendelson, E. (2009). *Introduction to mathematical logic*. CRC Press.

Prior, A. N. (1968). Egocentric logic. *Noûs, 178*, 191–207.

Seligman, J., Liu, F., & Girard, P. (2011). Logic in the community. In: Logic and its applications, (pp. 178–188). Springer.

Seligman, J., Liu, F., & Girard, P. (2013). Facebook and the epistemic logic of friendship. In: 14th conference on Theoretical Aspects of Rationality and Knowledge (TARK '13), January 2013, Chennai, India, (pp. 229–238).

Van Benthem, J., Girard, P., & Roy, O. (2009). Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic, 38*(1), 83–125.

Von Wright, G. H. (1963). *The logic of preference*. Edinburgh University Press.