Synthese (2023) 201:77
https://doi.org/10.1007/s11229-023-04060-2

ORIGINAL RESEARCH

An epistemic logic of preferences

Pavel Naumov[1] · **Anna Ovchinnikova**[2]

Received: 6 April 2022 / Accepted: 23 January 2023

Abstract
The article studies preferences of agents in a setting with imperfect information. For such a setting, the authors propose a new class of preferences. It is said that an agent prefers one statement over another if, among all indistinguishable worlds, the agent prefers the worlds where the first statement is true to those where the second one is true. The main technical result is a sound and complete logical system describing the interplay between a binary modality capturing preferences and the knowledge modality. The proof of completeness is using a newly proposed "tumbled pairs" technique.

Keywords Preference · Knowledge · Completeness · Axiomatisation · Ceteris paribus · Betterness

1 Introduction

In this article, we study the logical properties of the interplay between an agent's knowledge and preferences. As an example, consider an agent Alex who is headed towards a store to buy fruits. For the sake of this example, suppose that it is known that the store only sells three types of fruits: apples, bananas, and cantaloupes. Each of the fruits is sold either at a regular price or a fixed sale price. Alex's preferences are captured in Table 1.

For example, he prefers apples on sale over bananas on sale, but bananas on sale over apples at the regular price. Note that the regular price of his favourite fruit, the cantaloupe, is so high that at the regular price the cantaloupe becomes his last choice.

✉ Pavel Naumov
p.naumov@soton.ac.uk

Anna Ovchinnikova
oaa232001@mail.ru

[1] Agents, Interaction and Complexity Group, School of Electronics and Computer Science, University of Southampton, Southampton, UK

[2] School of Philosophy and Cultural Studies, National Research University Higher School of Economics, Moscow, Russia

Published online: 16 February 2023

Table 1 Alex's buying preferences in descending order

Preference	Fruit	Price
1	Cantaloupe	Sale
2	Apple	Sale
3	Banana	Sale
4	Apple	Regular
5	Banana	Regular
6	Cantaloupe	Regular

Before entering the store, Alex does not know what fruits are on sale, so he does not know yet which fruit he would prefer to buy. Suppose that, upon entering the store, Alex notices that cantaloupes are being sold at the regular price. Upon learning this information, he realises that he would rather buy apples, no matter if they are on sale or not, than a cantaloupe, see Table 1. We write this as:

$$\text{“Alex buys apples”} \rhd_{\text{Alex}} \text{“Alex buys a cantaloupe”}. \tag{1}$$

Similarly,

$$\text{“Alex buys bananas”} \rhd_{\text{Alex}} \text{“Alex buys a cantaloupe”}.$$

Note that, at this point, Alex still does not know the prices of apples and bananas. As a result, he does not yet have preferences between buying these two fruits, see Table 1:

$$\neg(\text{“Alex buys apples”} \rhd_{\text{Alex}} \text{“Alex buys bananas”}),$$
$$\neg(\text{“Alex buys bananas”} \rhd_{\text{Alex}} \text{“Alex buys apples”}).$$

Next, let us suppose that Alex goes deeper into the store and discovers that apples are on sale today. Based on this knowledge, he forms a preference to buy apples over bananas no matter if bananas are on sale or not, see Table 1:

$$\text{“Alex buys apples”} \rhd_{\text{Alex}} \text{“Alex buys bananas”}.$$

The goal of this work is to formally define and study the properties of *knowledge-informed preferences* expressed by modality $\rhd_a$.

Various logical systems for describing preferences have been proposed before. Åqvist (1962) introduces a three-valued logical system for "deontically better" modality. Chisholm and Sosa (1966) intuitively define "intrinsically better" as a modality, propose axioms for it and derive additional properties from these axioms. More recently, Van Benthem et al. (2009) consider "betterness" modality $[>]_a\varphi$ which denotes the fact that statement φ holds in all worlds that agent a prefers over the current world. Liu (2011, p.56) proposed a logical system combining modality $[>]$ with knowledge. Grossi et al. (2022) proposed several versions of "conditional best"

modality. Jiang and Naumov (2022) gave an axiomatisation of an egocentric modality "I prefer those agents who".

The most related to us and probably the most influential among earlier logical systems for preferences is Halldén's Logic of Better (1957). It captures "better" relation $p \rhd q$ between propositional variables in the perfect information setting. Halldén assumes that $p \rhd q$ if each world in which $p \wedge \neg q$ holds is at least as good as each world in which $\neg p \wedge q$ holds. A similar approach is also used later by Doyle et al. (1991). We discuss this type of preferences and compare it with our approach in Sect. 7.2.

There is a fundamental problem with Halldén's definition of $p \rhd q$ as "each world in which $p \wedge \neg q$ holds is at least as good as each world in which $\neg p \wedge q$ holds". Intuitively, sentence the "Alex prefers buying a cantaloupe over buying apples" does *not* mean that Alex prefers a world in which he buys a cantaloupe and has an incurable form of cancer to the world where he does not have cancer but has to settle on buying apples. To address this problem, Von Wright (1963) introduces *ceteris paribus* ("all other things being equal") principle. Instead of comparing all worlds, this principle requires comparing only the worlds where the other things are equal. For example, Alex prefers buying a cantaloupe to buying apples if, among the worlds where he has cancer, he prefers those where he buys a cantaloupe over those where he buys apples. The same for the worlds where he does not have cancer.

Although the ceteris paribus principle constitutes an important paradigm shift that significantly improves our understanding of preferences, it too has a problem. Namely, this principle depends on what exactly is "other things" that are being equal. Von Wright suggests that "other things" are all propositional variables in the language. This clarification of "other things" makes ceteris paribus too weak in some situations and too strong in others.

This clarification makes the principle too *weak* if the language of the model does not have enough propositional variables to distinguish worlds that should not be compared. To illustrate how this clarification makes the principle too *strong*, let us introduce a propositional variable r (receipt) that stands for "Alex's grocery receipt shows a cantaloupe". The ceteris paribus principle now forces us to compare the following two situations (worlds):

1. Alex buys a cantaloupe and his receipt shows this,
2. Alex buys apples and tricks the register into charging him for a cantaloupe.

Simultaneously, we also need to compare the situations:

1. Alex buys a cantaloupe and tricks the register into not charging him for it,
2. Alex buys apples and, as expected, his receipt does not show a cantaloupe.

This example shows that requiring all other propositional variables to be equal is perhaps too strong to capture what we intuitively mean by "Alex prefers buying a cantaloupe over buying apples".

An important next step in the formal treatment of ceteris paribus is done by Van Benthem et al. (2009). The consider modality $[\Gamma]_a\varphi$, where Γ is an arbitrary set of formulae. This modality stands for "statement φ is true in all worlds that agent a prefers over the current and in which all formulae in Γ have the same truth value as

in the current world". For example, imagine a world where Alex buys a cantaloupe at a regular price. In this world,

$$[\text{“cantaloupes are on sale”}]_{\text{Alex}}\text{“Alex buys apples”}. \tag{2}$$

Indeed, the statement "cantaloupes are on sale" is false in the current world and among all worlds where it is false, Alex prefers worlds where he buys apples to the current world where he buys a cantaloupe. As observed in Van Benthem et al. (2009), if set Γ is *finite* then, modality $[\Gamma]_a\varphi$ is expressible through discussed earlier "betterness" modality $[>]_a\varphi$. For example, statement (2) is equivalent to

$$\begin{aligned}&(\text{“cantaloupes are on sale”} \rightarrow\\ &\quad [>]_{\text{Alex}}(\text{“cantaloupes are on sale”} \rightarrow \text{“Alex buys apples”}))\\ &\wedge\\ &(\neg\text{“cantaloupes are on sale”} \rightarrow\\ &\quad [>]_{\text{Alex}}(\neg\text{“cantaloupes are on sale”} \rightarrow \text{“Alex buys apples”})).\end{aligned}$$

At the same time, if we allow Γ to be infinite, then the semantics of modality $[\Gamma]_a\varphi$ is problematic. Indeed, let Φ be the set of *all* formulae in our language and p be a propositional variable. Let φ be formula $[\Gamma]_a p$. Note that $\varphi \in \Gamma$. Thus, the satisfaction of formula φ is defined in terms of itself.

Recall that, in this work, we aim to study preferences informed by knowledge of the agent. Thus, such preferences would be the same in all worlds indistinguishable by the agents. In other words, such preferences could be described as *scitis paribus* ("all known being equal") preferences. Several possible modalities can be used to capture such preferences.

First, one can use Benthem, Girard, and Roy's modality $[\Gamma]_a p$ where Γ is the set of all true in the current world formulae of the form $K_a\psi$. This approach is problematic for at least two reasons. First, such set Γ would vary from world to world. Second, the described above problem with infinite Γ manifests itself here as well: can Γ include formula $K_a[\Gamma]_a p$?

Second, one can modify modality $[>]_a$ to mean "in all words better, to agent a, than the current world and indistinguishable from the current world by agent a". Note that it is easy to imagine a situation with two indistinguishable by agent a worlds such that the modified formula $[>]_a p$ is true in one of these worlds by not the other. Thus, such a modality captures preferences that might not be known to the agent. It is an interesting modality to study, but it is *not* the one that captures knowledge-informed preferences.

Finally, one can modify the original Halldén's relation $p \rhd q$ into modality $\varphi \rhd_a \psi$ that states that, among all worlds indistinguishable from the current world, agent a prefers worlds in which φ is true over worlds in which ψ is true. This is the approach adopted in the current paper. We propose a formal semantics of modality $\rhd$ and a sound and complete logical system that describes the interplay between this modality and the individual knowledge modality.

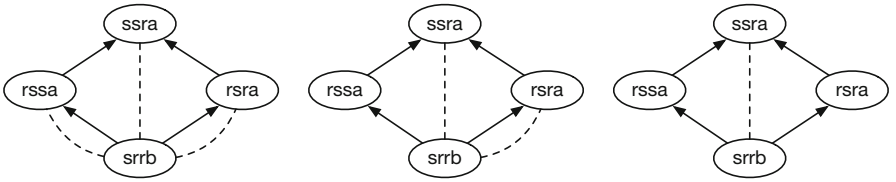

Fig. 1 Fragments of three epistemic models with preferences. Each diagram shows only 4 out of 24 epistemic worlds. The three models capture the knowledge of Alex at different moments

The rest of the article is structured as follows. First, we introduce epistemic models with preferences that are used in Sect. 3 to give a formal semantics of our logical system. We illustrate the formal semantics with additional examples in Sect. 4. Section 5 lists and discusses the axioms of our logical system. The soundness of these axioms is shown in the Proof of Soundeness section of the appendix. Section 6 contains our main technical result, the proof of completeness. In Sect. 7, we present additional properties of our system and its connection with other logical approaches to preferences. Section 8 concludes.

2 Epistemic model with preferences

In this section, we introduce epistemic models with preferences that are used later to give the formal semantics of preference modality $\rhd$. Throughout the article, we assume a fixed nonempty set of propositional variables and a fixed set of agents $\mathcal{A}$.

Definition 1 A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\succ_a\}_{a\in\mathcal{A}}, \pi)$ is an epistemic model with preferences, if

1. W is a set of "epistemic worlds",
2. $\sim_a$ is an "indistinguishability" equivalence relation on set W for each agent $a \in \mathcal{A}$,
3. $\succ_a$ is a "preference" strict partial order on set W for each agent $a \in \mathcal{A}$,
4. $\pi(p)$ is a subset of W for each propositional variable p.

For any sets $U, V \subseteq W$, we write $U \succ_a V$ if $u \succ_a v$ for all epistemic worlds $u \in U$ and $v \in V$. Note that item 3 above assumes that partial order $\succ_a$ is *strict*. We discuss this assumption in detail in Sect. 7.4.

In our introductory example, we considered knowledge of agent Alex at *three distinct moments*: before entering the store, after entering the store and discovering that cantaloupes are being sold at the regular price, and after going deeper into the store and learning that apples are on sale. As a result, to capture the introductory example we consider three epistemic models with preferences. These three models have the same set of possible worlds and the same preference relation, but different indistinguishability relations. Informally, these models capture "the mental state" of the agent at three distinct moments.

The epistemic worlds in all three models can be described as quadruples $w = abcf$, where $a, b, c \in$ {regular, sale} are the prices of apples, bananas, and cantaloupes, respectively, in world w and $f \in$ {apple, banana, cantaloupe} is the fruit that Alex buys

in world w. Note that we use more succinct notation $abcf$ for a tuple instead of more standard (a, b, c, f). There are $2 \times 2 \times 2 \times 3 = 24$ epistemic worlds in each model. The diagrams in Fig. 1 depict *fragments* of these three models. The nodes in the diagrams are labelled with tuples $abcf$ representing worlds. For example, nodes labelled with "rssa" in all three diagrams represent the world in which apples are sold at the regular (r) price, bananas and cantaloupes are on sale (s), and the agent decides to buy apples (a).

Note that a single epistemic world in our example captures a *sequence* of events: Alex enters the store, learns the price of the cantaloupes, proceeds deeper into the store, learns the price of the apples, and finally buys a fruit. Thus, metaphorically speaking, an epistemic world in our example is a "movie" rather than a static "snapshot".

The dashed lines in the diagrams in Fig. 1 represent indistinguishability relations. Since the first diagram depicts the situation before Alex enters the store, in this diagram, he cannot distinguish any of the worlds. The second diagram depicts the situation after he learns the price of cantaloupes. As a result, he now can distinguish the worlds in which the cantaloupe is on sale (like "rssa") from those where it is at the regular price (like "srrb"). The third diagram depicts the situation after he learns the price of apples. At this stage, he still cannot distinguish the worlds that differ only by the price of bananas, such as world "ssra" and "srra" (not shown). Also, note that he hasn't bought any fruits yet, so he cannot distinguish the worlds that differ only by the fruit that he buys, such as world "srra" and world "srrb".

The arrows in the diagrams in Fig. 1 show preference relations between worlds based on the information given in Table 1. For example, according to the table, the agent prefers buying apples at the regular price over buying bananas at the regular price. This is reflected in the diagram by the arrow from world "srrb", in which the agent buys bananas at the regular price, to world "rssa", in which the agent buys apples at the regular price. Note that the agent has no preferences between worlds "rssa" and "rsra" in those three diagrams. This is because in both worlds the agent buys apples at the regular price. The fact that the cantaloupe is on sale in one of these worlds is irrelevant because in both worlds he is buying apples, not a cantaloupe.

3 Syntax and semantics

In this section, we define the language of our formal system and the satisfaction relation between the worlds of an epistemic model with preferences and the formulae in this language.

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

$$\varphi := p \mid \neg\varphi \mid (\varphi \to \varphi) \mid K_a\varphi \mid (\varphi \rhd_a \varphi).$$

We follow the standard rules for the omission of the parentheses. We read formula $K_a\varphi$ as "agent a knows φ" and formula $\varphi \rhd_a \psi$ as "agent a prefers φ over ψ". We assume that conjunction $\wedge$, disjunction $\vee$, and constant false $\bot$ are defined in the standard way.

Next, we give a formal semantics of our preference modality $\rhd$. To understand the intuition behind this formal semantics, let us go back to our introductory example and

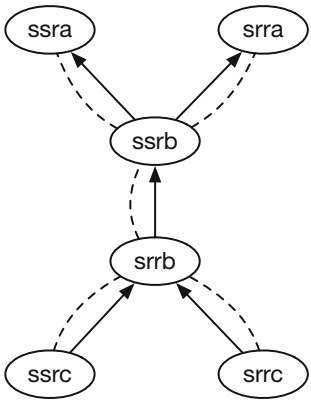

Fig. 2 Six worlds indistinguishable from "ssrc" after Alex learns that apples are on sale and cantaloupes are not

assume that the actual world is "ssrc". That is, apples and bananas are on sale and cantaloupes are not, but for whatever reason Alex ends up buying a cantaloupe. Recall that each of the three epistemic models with preferences has 24 possible worlds. Let us consider the model capturing the moment when Alex goes deeper into the store and discovers that apples are on sale. This model is partially depicted in the third diagram in Fig. 1.

In this model, there are only six worlds indistinguishable by Alex from the current world "ssrc". These worlds and the preference relations between them are depicted in Fig. 2. At the moment we consider, Alex knows that the current world is one of these six worlds, but he does not know which one. Note that, *among these six worlds, each world in which Alex buys apples is preferred to each world in which Alex buys bananas*. We interpret this as Alex's preference at that moment of buying apples over buying bananas:

$$\text{ssrc} \Vdash \text{"Alex buys apples"} \rhd_{\text{Alex}} \text{"Alex buys bananas"}.$$

This intuition is captured in item 5 of the definition below.

Definition 2 For any world $w \in W$ of an epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\succ_a\}_{a\in\mathcal{A}}, \pi)$ and any formula $\varphi \in \Phi$, the satisfaction relation $w \Vdash \varphi$ is defined recursively as follows:

1. $w \Vdash p$, if $w \in \pi(p)$,
2. $w \Vdash \neg\varphi$, if $w \nVdash \varphi$,
3. $w \Vdash \varphi \to \psi$, if $w \nVdash \varphi$ or $w \Vdash \psi$,
4. $w \Vdash K_a\varphi$, if $u \Vdash \varphi$ for each epistemic world $u \in W$ such that $w \sim_a u$,
5. $w \Vdash \varphi \rhd_a \psi$, when for all epistemic worlds $u, u' \in W$, if $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi$, and $u' \Vdash \psi$, then $u \succ_a u'$.

Item 5 states that in world w an agent prefers statement φ over statement ψ if, among all worlds indistinguishable from w, he prefers the worlds where φ is true to those where ψ is true. Note that for $\varphi \rhd_a \psi$ to hold, statements φ and ψ have to be mutually exclusive among indistinguishable worlds. We discuss an alternative approach in Sect. 7.2.

As with any approach, ours applies better to some settings than others. For example, in spite of our critique of the ceteris paribus principle in the introduction, it still has limited applicability to our introductory example. Some people will argue that the sentence "Alex prefers cantaloupe over apples" means that he prefers cantaloupe over apples when they are either both on sale or both not on sale. Our semantics of modality $\rhd$ in item 5 of Definition 2 does not model such interpretation of preferences.

On the other hand, our approach is good at modelling adaptive preferences, which are the preferences that change depending on the information available to the agent (Bruckner, 2009). For example, under our definition, once Alex learns that he does not get a gold medal, he starts preferring a silver one. Indeed, once Alex stops considering a world where he gets a gold medal as epistemically possible, among possible worlds he prefers those where he gets a silver medal to those where he does not. As a more controversial example, note that once Alex learns that he has a terminal disease, he prefers to die over staying alive. Indeed, in this situation he vacuously prefers each world where he dies to each element of the empty set of the worlds in which he stays alive.

It is easy to see that knowledge modality K_a is definable through preferences modality $\rhd_a$ as follows: $K_a\varphi \equiv (\neg\varphi) \rhd_a \neg\varphi$. Instead of proving this semantically in the current section, we derive the same property from the axioms of our logical system in Theorem 3 (Sect. 7.1). In spite of the definability of modality K_a through modality $\rhd_a$, we have chosen to keep them both as primary modalities in language Φ to improve the readability of the axioms.

To illustrate Definition 2, below we formally prove statement (1) about our introductory example.

Observation 1 *For any* $x, y \in \{r, s\}$ *and any* $f \in \{a, b, c\}$,

$$xyrf \Vdash \text{“Alex buys apples”} \rhd_{Alex} \text{“Alex buys a cantaloupe”}.$$

Proof Consider worlds $x'y'z'f'$ and $x''y''z''f''$ such that $xyrf \sim_{\text{Alex}} x'y'z'f'$ and $xyrf \sim_{\text{Alex}} x''y''z''f''$. Suppose that $x'y'z'f' \Vdash$ “Alex buys apples” and $x''y''z''f'' \Vdash$ “Alex buys a cantaloupe”. By item 5 of Definition 2, it suffices to show that $x'y'z'f' \succ_{\text{Alex}} x''y''z''f''$.

Recall that we consider the moment when Alex already knows the price of the cantaloupes. Thus, $z' = z'' = r$ by the assumptions $xyrf \sim_{\text{Alex}} x'y'z'f'$ and $xyrf \sim_{\text{Alex}} x''y''z''f''$. Also, $f' = a$ and $f'' = c$ by the assumptions $x'y'z'f' \Vdash$ “Alex buys apples” and $x''y''z''f'' \Vdash$ “Alex buys a cantaloupe”, respectively. Hence, it suffices to show that $x'y'ra \succ_{\text{Alex}} x''y''rc$. The last statement is true because Alex prefers buying apples at any price x' over buying a cantaloupe at the regular price, see Table 1. □

4 Additional examples

We further illustrate our formal definitions using the classical Prisoner's Dilemma and Battle of the Sexes games (Osborne and Rubinstein, 1994, p.15–16). The first of them

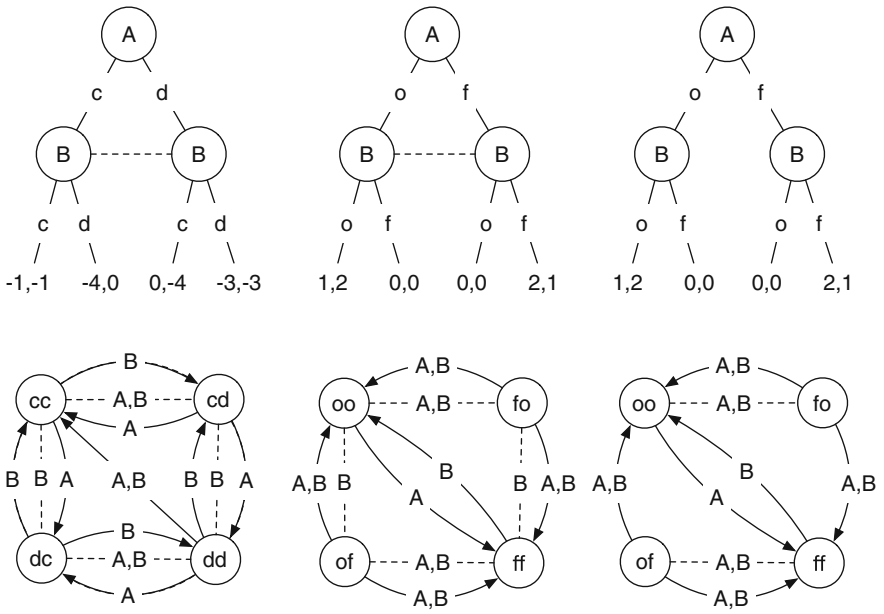

Fig. 3 Three extensive form games: Prisoner's Dilemma (left), Battle of the Sexes with imperfect information (centre), and Battle of the Sexes with perfect information (right)

is depicted in the extensive form in the upper left diagram in Fig. 3. Here, Alex (A) first chooses to cooperate (c) or to defect (d). Then, Brittany (B) also chooses to cooperate or to defect. We assume that this is an extensive form game with imperfect information in the sense that Brittany does not know Alex's choice when she makes her own. We capture this by the assumption that the two states labelled with B in the diagram are indistinguishable by Brittany and visualise this by the dashed line connecting these two states.

The lower-left diagram in Fig. 3 depicts "ex interim" epistemic model with preferences describing the Prisoner's Dilemma game. This model has four possible epistemic worlds corresponding to four possible paths of play in the game. For example, the world labelled with dc corresponds to the scenario in which Alex defects and Brittany cooperates. The indistinguishability relations (shown using dashed lines) represent the knowledge of the agents *after* Alex has made his choice and *before* Brittany has made hers. Recall that she makes her choice not knowing Alex's choice. Thus, for example, Alex can distinguish world cc, in which he cooperates, from world dc, in which he defects, but Brittany cannot.

The diagram in the upper-left corner also shows Alex's and Brittany's utility functions (negative number of years spent in prison). For example, if Alex cooperates and Brittany defects, then Alex spends four years in prison (utility -4) and Brittany is released immediately (utility 0). We assume that Alex and Brittany are rational agents – they prefer the worlds where their utility function is higher to those worlds where it is lower. For example, Alex prefers world cc, where his utility is -1, to world cd,

where his utility is -4. In the lower-left diagram, we show preference relations using labelled arrows. For example, an arrow from world cd to world cc labelled with A means that Alex prefers world cc to world cd.

Let us now consider world cc. Recall that we discuss the moment when Alex has already made his decision to cooperate. Brittany does not know which decision he has made and she has not yet decided on her own move. At that moment, Alex considers only two worlds to be possible: cc and cd. He prefers the first over the second, see Fig. 3 (lower left). Brittany cooperates in world cc and she defects in world cd. Thus, by item 5 of Definition 2,

$$cc \Vdash \text{“Brittany cooperates”} \rhd_{\text{Alex}} \text{“Brittany defects”}. \tag{3}$$

Since Brittany does not know Alex's move and she has not yet decided on her own, she considers all four worlds of this game cc, cd, dc, and dd to be possible. Note that she prefers each of the worlds in the set $\{cd, dd\}$ to each of the worlds in the set $\{cc, dc\}$, see Fig. 3 (lower left). Brittany defects in each of the worlds from the first set and cooperates in each of the worlds from the second set. Thus, by item 5 of Definition 2,

$$cc \Vdash \text{“Brittany defects”} \rhd_{\text{Brittany}} \text{“Brittany cooperates”}. \tag{4}$$

Our second example, the Battle of the Sexes game with imperfect information, is depicted in the middle of Fig. 3. Consider world oo where, choosing between opera (o) and football (f), Alex and Brittany both eventually decide to go to opera. We again consider the moment when Alex has already made his choice, but Brittany does not know it and she has not made her decision yet. At that moment, Alex considers only worlds oo and of to be possible. He prefers the first over the second. Brittany goes to opera in the first and to football in the second. Thus, by item 5 of Definition 2, similarly to (3),

$$oo \Vdash \text{“Brittany goes to opera”} \rhd_{\text{Alex}} \text{“Brittany goes to football”}.$$

Note, however, that an equivalent of statement (4) does not hold in the Battle of the Sexes game. Namely,

$$oo \nVdash \text{“Brittany goes to opera”} \rhd_{\text{Brittany}} \text{“Brittany goes to football”}.$$

This is because at that moment, Brittany allows the possibility of all four worlds and, for example, she does *not* prefer world fo to world ff, see Fig. 3 (lower middle).

Finally, note that in the perfect information version of the same game, see Fig. 3 (right), Brittany knows Alex's choice to go to opera. Thus, she only considers worlds oo and of to be possible. She prefers the former over the latter. Thus,

$$oo \Vdash \text{“Brittany goes to opera”} \rhd_{\text{Brittany}} \text{“Brittany goes to football”}.$$

In other words, the extra information available to Brittany in the perfect information setting allows her to form the preference to go to opera rather than to football.

In conclusion, note that preferences and knowledge modalities can be combined to express additional properties. For example, in the Prisoner's Dilemma example, in any world,

$$(K_{\text{Alex}}(\text{“Alex will not be released immediately”})) \rhd_{Brittany} (K_{\text{Alex}}(\text{“Brittany will not be released immediately”})). \quad (5)$$

Indeed, the statement $K_{\text{Alex}}(\text{“Alex will not be released immediately”})$ is satisfied in worlds cc and cd. At the same time, it can also be observed that the statement $K_{\text{Alex}}(\text{“Brittany will not be released immediately”})$ is satisfied in worlds dc and dd. Thus, statement (5) holds because Brittany prefers each of the worlds in the set $\{cc, cd\}$ to each of the worlds in the set $\{dc, dd\}$, see Fig. 3 (lower left).

5 Axioms

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

Truth	$K_a\varphi \to \varphi$
Negative Introspection	$\neg K_a\varphi \to K_a\neg K_a\varphi$
Distributivity	$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$
Introspection of Preference	$\varphi \rhd_a \psi \to K_a(\varphi \rhd_a \psi)$
Monotonicity	$K_a(\varphi \to \psi) \to (\psi \rhd_a \chi \to \varphi \rhd_a \chi)$
	$K_a(\varphi \to \psi) \to (\chi \rhd_a \psi \to \chi \rhd_a \varphi)$
Strictness	$\varphi \rhd_a \psi \to \neg(\varphi \wedge \psi)$
Superiority of Falsehood	$\bot \rhd_a \varphi$
Inferiority of Falsehood	$\varphi \rhd_a \bot$
Transitivity	$\varphi \rhd_a \psi \to (\psi \to (\psi \rhd_a \chi \to \varphi \rhd_a \chi))$
Disjunction	$\varphi \rhd_a \chi \to (\psi \rhd_a \chi \to (\varphi \vee \psi) \rhd_a \chi)$
	$\varphi \rhd_a \psi \to (\varphi \rhd_a \chi \to \varphi \rhd_a (\psi \vee \chi))$

The Truth, the Negative Introspection, and the Distributivity are the standard axioms of the epistemic logic S5. The Introspection of Preference axiom states that each agent knows his preferences. Of course, in general, this is not true. However, this is true for the class of "all other known being equal" preferences defined by item 5 of Definition 2. The two forms of the Monotonicity axiom together say that each side of a preference claim can be replaced by a knowingly-stronger statement.

The Strictness axiom states that if an agent prefers φ over ψ, then one of these statements must be false in the current world w. The axiom is true because the assumption of the opposite implies that $w \prec_a w$ by item 5 of Definition 2. The latter contradicts relation $\prec_a$ being a strict order. The Strictness axiom is also valid in a stronger form

$\varphi \rhd_a \psi \rightarrow K_a \neg(\varphi \wedge \psi)$, which is provable in our logical system. We further discuss the strictness assumption in Sect. 7.4.

The Superiority of Falsehood axiom states that an agent prefers a world where statement $\bot$ is true to any world in the model. The Inferiority of Falsehood axiom states that an agent prefers any world in the model to a world where statement $\bot$ is true. These axioms are vacuously true because statement $\bot$ is false in all worlds.

The Transitivity axiom captures the fact that preference relation on epistemic worlds is transitive. The second assumption of this axiom, formula ψ, is significant. In the form without this assumption: $\varphi \rhd_a \psi \rightarrow (\psi \rhd_a \chi \rightarrow \varphi \rhd_a \chi)$ the axiom is not valid. Indeed, in the case $\psi \equiv \bot$, this hypothetical axiom has the form: $\varphi \rhd_a \bot \rightarrow (\bot \rhd_a \chi \rightarrow \varphi \rhd_a \chi)$. Note that $\varphi \rhd_a \bot$ and $\bot \rhd_a \chi$ are instances of the Inferiority of Falsehood and the Superiority of Falsehood axioms, respectively. Thus, the hypothetical axiom implies that $\varphi \rhd_a \chi$ for arbitrary formulae φ and χ. The Transitivity axiom (with the second assumption) is valid in a more general form $\varphi \rhd_a \psi \rightarrow (\neg K_a \neg \psi \rightarrow (\psi \rhd_a \chi \rightarrow \varphi \rhd_a \chi))$, which is provable in our logical system.

The first Disjunction axiom states that if an agent prefers the worlds where φ is true to those where χ is true and he also prefers the worlds where ψ is true to those where χ is true, then he prefers the worlds where $\varphi \vee \chi$ is true to those where χ is true. The second Disjunction axiom states a similar principle for the right-hand side of the modality $\rhd$.

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

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi} \qquad \frac{\varphi}{K_a \varphi}.$$

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

The next strong soundness theorem is proven in the Proof of Soundness section of the appendix.

Theorem 1 *For any world w of an epistemic model with preferences, any set of formulae $X \subseteq \Phi$, and any formula $\varphi \in \Phi$ if $w \Vdash \chi$ for each formula $\chi \in X$ and $X \vdash \varphi$, then $w \Vdash \varphi$.*

6 Completeness

In this section, we prove the completeness of our logical system. We start this proof with Sect. 6.1, which states several auxiliary lemmas about derivability in our system. Section 6.2 defines the canonical epistemic model with preferences. In the section that follows, we discuss and formally define the notion of a tumbled pair of sets of

formulae. Finally, in Sect. 6.4, we use tumbled pairs and the canonical model to prove the strong completeness.

6.1 Auxiliary lemmas

Lemma 1 *1.* $\vdash \varphi_1 \rhd_a \psi_1 \to (\varphi_2 \rhd_a \psi_2 \to (\varphi_1 \vee \varphi_2) \rhd_a (\psi_1 \wedge \psi_2))$,
2. $\vdash \varphi_1 \rhd_a \psi_1 \to (\varphi_2 \rhd_a \psi_2 \to (\varphi_1 \wedge \varphi_2) \rhd_a (\psi_1 \vee \psi_2))$.

Proof To prove the first of these statements, note that formulae $\psi_1 \wedge \psi_2 \to \psi_1$ and $\psi_1 \wedge \psi_2 \to \psi_2$ are propositional tautologies. Thus, by the Necessitation inference rule, $\vdash K_a(\psi_1 \wedge \psi_2 \to \psi_1)$ and $\vdash K_a(\psi_1 \wedge \psi_2 \to \psi_2)$. Hence, by the second Monotonicity axiom and the Modus Ponens inference rule,

$$\vdash \varphi_1 \rhd \psi_1 \to \varphi_1 \rhd (\psi_1 \wedge \psi_2),$$
$$\vdash \varphi_2 \rhd \psi_2 \to \varphi_2 \rhd (\psi_1 \wedge \psi_2).$$

At the same time, the following formula is an instance of the first Disjunction axiom:

$$\varphi_1 \rhd (\psi_1 \wedge \psi_2) \to (\varphi_2 \rhd (\psi_1 \wedge \psi_2) \to (\varphi_1 \vee \varphi_2) \rhd (\psi_1 \wedge \psi_2)).$$

Therefore, by propositional reasoning,

$$\vdash \varphi_1 \rhd_a \psi_1 \to (\varphi_2 \rhd_a \psi_2 \to (\varphi_1 \vee \varphi_2) \rhd_a (\psi_1 \wedge \psi_2)).$$

The proof of the other statement is similar, using the first Monotonicity axiom and the second Disjunction axiom. □

Lemma 2 $(\chi \wedge \neg\varphi) \rhd_a \gamma, \varphi \rhd_a \psi, \chi \rhd_a (\gamma \wedge \neg\psi) \vdash \chi \rhd_a \gamma$.

Proof By item 1 of Lemma 1,

$$\vdash (\chi \wedge \neg\varphi) \rhd_a \gamma \to (\varphi \rhd_a \psi \to ((\chi \wedge \neg\varphi) \vee \varphi) \rhd_a (\gamma \wedge \psi)).$$

At the same time, by item 2 of Lemma 1,

$$\vdash ((\chi \wedge \neg\varphi) \vee \varphi) \rhd_a (\gamma \wedge \psi) \to (\chi \rhd_a (\gamma \wedge \neg\psi) \to (((\chi \wedge \neg\varphi) \vee \varphi) \wedge \chi) \rhd_a ((\gamma \wedge \psi) \vee (\gamma \wedge \neg\psi))).$$

Thus, from the two previous statements by the Modus Ponens inference rule applied four times,

$$(\chi \wedge \neg\varphi) \rhd_a \gamma, \varphi \rhd_a \psi, \chi \rhd_a (\gamma \wedge \neg\psi) \vdash (((\chi \wedge \neg\varphi) \vee \varphi) \wedge \chi) \rhd_a ((\gamma \wedge \psi) \vee (\gamma \wedge \neg\psi)).$$

Note that formula $\chi \rightarrow ((\chi \wedge \neg\varphi) \vee \varphi) \wedge \chi$ is a propositional tautology. Hence, $\vdash K_a(\chi \rightarrow ((\chi \wedge \neg\varphi) \vee \varphi) \wedge \chi)$ by the Necessitation inference rule. Then, by the first Monotonicity axiom and the Modus Ponens inference rule,

$$(\chi \wedge \neg\varphi) \rhd_a \gamma, \varphi \rhd_a \psi, \chi \rhd_a (\gamma \wedge \neg\psi) \vdash \chi \rhd_a ((\gamma \wedge \psi) \vee (\gamma \wedge \neg\psi)).$$

Next, observe that formula $\gamma \rightarrow (\gamma \wedge \psi) \vee (\gamma \wedge \neg\psi)$ is also a propositional tautology. Thus, $\vdash K_a(\gamma \rightarrow (\gamma \wedge \psi) \vee (\gamma \wedge \neg\psi))$ by the Necessitation inference rule. Therefore,

$$(\chi \wedge \neg\varphi) \rhd_a \gamma, \varphi \rhd_a \psi, \chi \rhd_a (\gamma \wedge \neg\psi) \vdash \chi \rhd_a \gamma$$

by the second Monotonicity axiom and the Modus Ponens inference rule. □

Lemma 3 *(Lindenbaum) Any consistent set of formulae can be extended to a maximal consistent set of formulae.*

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

We omit the proofs of the next three well-known lemmas.

Lemma 4 *(positive introspection)* $\vdash K_a\varphi \rightarrow K_aK_a\varphi$.

Lemma 5 *(deduction) If* $X, \varphi \vdash \psi$, *then* $X \vdash \varphi \rightarrow \psi$.

Lemma 6 *If* $\varphi_1, \ldots, \varphi_n \vdash \psi$, *then* $K_a\varphi_1, \ldots, K_a\varphi_n \vdash K_a\psi$.

6.2 Canonical model

In this section, we proceed to define the canonical epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\succ_a\}_{a\in\mathcal{A}}, \pi)$. Like in most proofs of completeness for modal logics, the epistemic worlds of the canonical model are maximal consistent sets of formulae.

Definition 3 W is the set of all maximal consistent sets of formulae.

Definition 4 For all worlds $w, u \in W$, let $w \sim_a u$ when for each formula φ, if $K_a\varphi \in w$, then $\varphi \in u$.

One can also define that $w \sim_a u$ is true if maximal consistent sets of formulae w and u have the same K_a-formulae. Although our definition results in shorter proofs of several lemmas, it is a bit cumbersome to show that the relation $\sim_a$, as it is defined above, is an equivalence relation. We give the proof of the lemma below in the Auxiliary Lemma section of the appendix.

Lemma 7 *Relation* $\sim_a$ *is an equivalence relation on set W for each* $a \in \mathcal{A}$.

The next definition is a straightforward reflection of item 5 of Definition 2.

Definition 5 $w \succ_a u$ if $w \sim_a u$ and there is a formula $\varphi \rhd_a \psi \in w$ such that $\varphi \in w$ and $\psi \in u$.

Next, we show that relation $\succ_a$ is a strict partial order. Note that the fact that it is transitive is non-trivial and, at least to us, unexpected. Indeed, the assumptions $w \succ_a u$ and $u \succ_a v$ imply that there are formula $\varphi_1 \rhd_a \psi_1 \in w$ and formula $\varphi_2 \rhd_a \psi_2 \in u$ such that $\varphi_1 \in w$, $\psi_1 \in u$, $\varphi_2 \in u$, and $\psi_2 \in v$. It is not obvious how this would imply that $w \succ_a v$.

Lemma 8 *Relation $\succ_a$ is a strict partial order on set W.*

Proof **Irreflexivity.** Suppose that $w \succ_a w$ for some epistemic world $w \in W$. Thus, by Definition 5, there exists a formula $\varphi \rhd_a \psi \in w$ such that $\varphi, \psi \in w$. The statement $\varphi \rhd_a \psi \in w$ implies $w \vdash \neg(\varphi \wedge \psi)$ by the Strictness axiom and the Modus Ponens inference rule. At the same time, the statement $\varphi, \psi \in w$ implies $w \vdash \varphi \wedge \psi$ by the laws of propositional reasoning. Therefore, set w is not consistent, which contradicts Definition 3.

Transitivity. Suppose that $w \succ_a u$ and $u \succ_a v$. Thus, by Definition 5,

$$w \sim_a u \text{ and } u \sim_a v \tag{6}$$

and there are formulae

$$\varphi_1 \rhd_a \psi_1 \in w \text{ and } \varphi_2 \rhd_a \psi_2 \in u \tag{7}$$

such that

$$\varphi_1 \in w, \psi_1 \in u \text{ and } \varphi_2 \in u, \psi_2 \in v. \tag{8}$$

Note that $\psi_1 \wedge \varphi_2 \rightarrow \psi_1$ and $\psi_1 \wedge \varphi_2 \rightarrow \varphi_2$ are propositional tautologies. Thus, $\vdash K_a(\psi_1 \wedge \varphi_2 \rightarrow \psi_1)$ and $\vdash K_a(\psi_1 \wedge \varphi_2 \rightarrow \varphi_2)$ by the Necessitation inference rule. Hence, by the Monotonicity axiom and the Modus Ponens inference rule using statement (7),

$$w \vdash \varphi_1 \rhd_a (\psi_1 \wedge \varphi_2) \text{ and } u \vdash (\psi_1 \wedge \varphi_2) \rhd_a \psi_2. \tag{9}$$

The first conjunct in the above formula implies $w \vdash K_a(\varphi_1 \rhd_a (\psi_1 \wedge \varphi_2))$ by the Introspection of Preference axiom and the Modus Ponens inference rule. Thus, $K_a(\varphi_1 \rhd_a (\psi_1 \wedge \varphi_2)) \in w$ because set w is maximal. Hence, $\varphi_1 \rhd_a (\psi_1 \wedge \varphi_2) \in u$ by Definition 4 and the part $w \sim_a u$ of statement (6). Then, by the Transitivity axiom and the Modus Ponens inference rule,

$$u \vdash \psi_1 \wedge \varphi_2 \rightarrow ((\psi_1 \wedge \varphi_2) \rhd_a \psi_2 \rightarrow \varphi_1 \rhd_a \psi_2).$$

Thus, by propositional reasoning using the parts $\psi_1 \in u$ and $\varphi_2 \in u$ of statement (8),

$$u \vdash (\psi_1 \wedge \varphi_2) \rhd_a \psi_2 \rightarrow \varphi_1 \rhd_a \psi_2.$$

Hence, $u \vdash \varphi_1 \rhd_a \psi_2$ by the Modus Ponens inference rule and the second conjunct from statement (9). Then, $\varphi_1 \rhd_a \psi_2 \in u$ because set u is maximal. At the same time,

$w \sim_a v$ by Lemma 7 and statement (6). Also, $\varphi_1 \in w$ and $\psi_2 \in v$ by statement (8). Therefore, $w \succ_a v$ by Definition 5. □

Definition 6 $\pi(p) = \{w \in W \mid p \in w\}$ for each propositional variable p.

This concludes the definition of the canonical epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\succ_a\}_{a\in\mathcal{A}}, \pi)$.

6.3 Tumbled pairs

As usual, at the core of the proof of completeness is an "induction" or a "truth" lemma. In our case, this is Lemma 17. The proof of the induction lemma in modal logic is often preceded by one or more lemmas that show that if a maximal consistent set of formulae w does *not* contain a modal formula $\Box\varphi$, then there is a maximal consistent set u, somehow related to set w, such that $\varphi \notin u$. In our case, for modality K_a, this is Lemma 14. The situation is more complicated in the case of modality $\rhd_a$. Indeed, the corresponding lemma for modality $\rhd_a$, Lemma 15, claims the existence of not one but two interdependent maximal consistent sets because item 5 of Definition 2 refers to two worlds, u and u'. To prove Lemma 15, one needs to construct *simultaneously* these two sets. Towards this goal, we have developed a "tumbled pairs" technique that we describe in this section. We use this technique in the proof of Lemma 15.

We say that two (not necessarily maximal consistent) sets of formulae form a "tumbled pair" if they satisfy a certain constraint specified in Definition 7 below. Technically, we talk about a (w, a)-tumbled pair because the constraint depends on parameters w and a. In Lemma 11, we show that a specific pair is tumbled. In Lemma 12, we show that any tumbled pair can be extended in a certain way while still remaining tumbled. If such an extension is performed *ad infinitum*, then we say that the tumbled pair is *saturated*, see Definition 8. In Lemma 13, we observe that any tumbled pair can be extended to a saturated tumbled pair. In the next section, to prove Lemma 15, we start with the tumbled pair from Lemma 11, extend it to a saturated tumbled pair by Lemma 13, and, finally, further extend it to a pair of maximal consistent sets of formulae u and u' using the Lindenbaum's lemma.

Definition 7 Let w be a maximal consistent set of formulae, $a \in \mathcal{A}$ be an agent, and X, Y be two (possibly infinite) sets of formulae. Pair (X, Y) is (w, a)-tumbled if $(\wedge X') \rhd_a (\wedge Y') \notin w$ for any finite sets $X' \subseteq X$ and $Y' \subseteq Y$.

Lemma 9 *If pair (X, Y) is (w, a)-tumbled, then set X is consistent.*

Proof If set X is inconsistent, then there must exist a finite set $X' \subseteq X$ such that $X' \vdash \bot$. Hence, $\vdash \wedge X' \to \bot$ by Lemma 5 and propositional reasoning. Thus, $\vdash K_a(\wedge X' \to \bot)$ by the Necessitation inference rule. Let set Y' be an arbitrary finite subset of Y. Then, by the first Monotonicity axiom and the Modus Ponens inference rule,

$$\vdash (\bot \rhd_a \wedge Y') \to (\wedge X' \rhd_a \wedge Y').$$

Note that formula $\bot \rhd_a \wedge Y'$ is an instance of the Superiority of Falsehood axiom. Hence, by the Modus Ponens inference rule,

$$\vdash \wedge X' \rhd_a \wedge Y'.$$

Then, $\wedge X' \rhd_a \wedge Y' \in w$ because set w is maximal. Therefore, pair (X, Y) is not (w, a)-tumbled by Definition 7. □

Lemma 10 *If pair (X, Y) is (w, a)-tumbled, then set Y is consistent.*

Proof If set Y is inconsistent, then there must exist a finite set $Y' \subseteq Y$ such that $Y' \vdash \bot$. Hence, $\vdash \wedge Y' \rightarrow \bot$ by Lemma 5 and propositional reasoning. Thus, $\vdash K_a(\wedge Y' \rightarrow \bot)$ by the Necessitation inference rule. Let set X' be an arbitrary finite subset of X. Thus, by the second Monotonicity axiom and the Modus Ponens inference rule,

$$\vdash (\wedge X' \rhd_a \bot) \rightarrow (\wedge X' \rhd_a \wedge Y').$$

Note that formula $\wedge X' \rhd_a \bot$ is an instance of the Inferiority of Falsehood axiom. Hence, by the Modus Ponens inference rule,

$$\vdash \wedge X' \rhd_a \wedge Y'.$$

Then, $\wedge X' \rhd_a \wedge Y' \in w$ because set w is maximal. Therefore, pair (X, Y) is not (w, a)-tumbled by Definition 7. □

Lemma 11 *For any maximal consistent set of formulae w and any formula $\varphi \rhd_a \psi \notin w$, pair (X, Y) is (w, a)-tumbled, where*

$$X = \{\varphi\} \cup \{\chi \mid K_a\chi \in w\},$$
$$Y = \{\psi\} \cup \{\gamma \mid K_a\gamma \in w\}.$$

Proof Suppose the opposite. Thus, by Definition 7, there are finite sets X' and Y' such that

$$X' \subseteq \{\varphi\} \cup \{\chi \mid K_a\chi \in w\}, \tag{10}$$
$$Y' \subseteq \{\psi\} \cup \{\gamma \mid K_a\gamma \in w\}, \tag{11}$$
$$\wedge X' \rhd_a \wedge Y' \in w. \tag{12}$$

Assume that $\chi_1, \ldots, \chi_m$ is the list of all the formulae in the finite set $X' \cap \{\chi \mid K_a\chi \in w\}$ and $\gamma_1, \ldots, \gamma_n$ is the list of all the formulae in the finite set $Y' \cap \{\gamma \mid K_a\gamma \in w\}$. Then, by statements (10) and (11),

$$\chi_1, \ldots, \chi_m, \varphi \vdash \wedge X',$$
$$\gamma_1, \ldots, \gamma_n, \psi \vdash \wedge Y'.$$

Hence, by Lemma 5,

$$\chi_1, \ldots, \chi_m \vdash \varphi \to \wedge X',$$
$$\gamma_1, \ldots, \gamma_n \vdash \psi \to \wedge Y'.$$

Thus, by Lemma 6,

$$K_a\chi_1, \ldots, K_a\chi_m \vdash K_a(\varphi \to \wedge X'),$$
$$K_a\gamma_1, \ldots, K_a\gamma_n \vdash K_a(\psi \to \wedge Y').$$

Recall that $K_a\chi_1, \ldots, K_a\chi_m, K_a\gamma_1, \ldots, K_a\gamma_n \in w$ by the choice of formulae $\chi_1, \ldots, \chi_m, \gamma_1, \ldots, \gamma_n$. Then,

$$w \vdash K_a(\varphi \to \wedge X'),$$
$$w \vdash K_a(\psi \to \wedge Y').$$

Hence, by the first and the second Monotonicity axioms and the Modus Ponens inference rule,

$$w \vdash (\wedge X' \rhd_a \wedge Y') \to (\varphi \rhd_a \wedge Y'),$$
$$w \vdash (\varphi \rhd_a \wedge Y') \to (\varphi \rhd_a \psi).$$

Thus, by the laws of propositional reasoning,

$$w \vdash (\wedge X' \rhd_a \wedge Y') \to (\varphi \rhd_a \psi).$$

Then, $w \vdash \varphi \rhd_a \psi$ by the Modus Ponens inference rule and statement (12). Therefore, $\varphi \rhd_a \psi \in w$ because set w is maximal, which contradicts the assumption $\varphi \rhd_a \psi \notin w$ of the lemma. □

Lemma 12 *For any (w, a)-tumbled pair (X, Y) and any formula $\varphi \rhd_a \psi \in w$, either pair $(X \cup \{\neg\varphi\}, Y)$ or pair $(X, Y \cup \{\neg\psi\})$ is (w, a)-tumbled.*

Proof Suppose pairs $(X \cup \{\neg\varphi\}, Y)$ and $(X, Y \cup \{\neg\psi\})$ are not (w, a)-tumbled. Thus, by Definition 7, there exist finite sets $X_1 \subseteq X \cup \{\neg\varphi\}$ and $Y_1 \subseteq Y$ such that

$$\wedge X_1 \rhd_a \wedge Y_1 \in w \tag{13}$$

and finite sets $X_2 \subseteq X$ and $Y_2 \subseteq Y \cup \{\neg\psi\}$ such that

$$\wedge X_2 \rhd_a \wedge Y_2 \in w. \tag{14}$$

The statements $X_1 \subseteq X \cup \{\neg\varphi\}$ and $X_2 \subseteq X$ imply that there is a finite set $X_0 \subseteq X$ such that

$$X_0, \neg\varphi \vdash \wedge X_1, \tag{15}$$

$$X_0 \vdash \wedge X_2. \tag{16}$$

Similarly, the statements $Y_1 \subseteq Y$ and $Y_2 \subseteq Y \cup \{\neg\psi\}$ imply that there is a finite set $Y_0 \subseteq Y$ such that

$$Y_0 \vdash \wedge Y_1, \tag{17}$$
$$Y_0, \neg\psi \vdash \wedge Y_2. \tag{18}$$

From statements (15), (16), (17), and (18) using Lemma 5 and propositional reasoning,

$$\begin{aligned}
&\vdash (\wedge X_0) \wedge \neg\varphi \rightarrow \wedge X_1,\\
&\vdash \wedge X_0 \rightarrow \wedge X_2,\\
&\vdash \wedge Y_0 \rightarrow \wedge Y_1,\\
&\vdash (\wedge Y_0) \wedge \neg\psi \rightarrow \wedge Y_2.
\end{aligned}$$

Hence, by the Necessitation inference rule,

$$\begin{aligned}
&\vdash K_a((\wedge X_0) \wedge \neg\varphi \rightarrow \wedge X_1),\\
&\vdash K_a(\wedge X_0 \rightarrow \wedge X_2),\\
&\vdash K_a(\wedge Y_0 \rightarrow \wedge Y_1),\\
&\vdash K_a((\wedge Y_0) \wedge \neg\psi \rightarrow \wedge Y_2).
\end{aligned}$$

Then, by the Monotonicity axioms and the Modus Ponens inference rule,

$$\begin{aligned}
&\vdash (\wedge X_1 \rhd_a \wedge Y_1) \rightarrow (((\wedge X_0) \wedge \neg\varphi) \rhd_a \wedge Y_1),\\
&\vdash (((\wedge X_0) \wedge \neg\varphi) \rhd_a \wedge Y_1) \rightarrow (((\wedge X_0) \wedge \neg\varphi) \rhd_a \wedge Y_0),\\
&\vdash (\wedge X_2 \rhd_a \wedge Y_2) \rightarrow (\wedge X_0 \rhd_a \wedge Y_2),\\
&\vdash (\wedge X_0 \rhd_a \wedge Y_2) \rightarrow (\wedge X_0 \rhd_a ((\wedge Y_0) \wedge \neg\psi)).
\end{aligned}$$

Thus, by the laws of propositional reasoning,

$$\begin{aligned}
&\vdash (\wedge X_1 \rhd_a \wedge Y_1) \rightarrow (((\wedge X_0) \wedge \neg\varphi) \rhd_a \wedge Y_0),\\
&\vdash (\wedge X_2 \rhd_a \wedge Y_2) \rightarrow (\wedge X_0 \rhd_a ((\wedge Y_0) \wedge \neg\psi)).
\end{aligned}$$

Hence, by the Modus Ponens inference rule using assumptions (13) and (14),

$$\begin{aligned}
&w \vdash ((\wedge X_0) \wedge \neg\varphi) \rhd_a (\wedge Y_0),\\
&w \vdash (\wedge X_0) \rhd_a ((\wedge Y_0) \wedge \neg\psi).
\end{aligned}$$

Then, by Lemma 2 using the assumption $\varphi \rhd_a \psi \in w$ of the lemma,

$$w \vdash \wedge X_0 \rhd_a \wedge Y_0.$$

Thus, $\wedge X_0 \rhd_a \wedge Y_0 \in w$ because set w is maximal. Therefore, by Definition 7, pair (X, Y) is not (w, a)-tumbled. □

Definition 8 A (w, a)-tumbled pair (X, Y) is (w, a)-saturated if for each formula $\varphi \rhd_a \psi \in w$, either $\neg\varphi \in X$ or $\neg\psi \in Y$.

The next lemma follows from Lemma 12 and Definition 8.

Lemma 13 *For any (w, a)-tumbled pair (X, Y), there is a (w, a)-saturated (w, a)-tumbled pair (X', Y') such that $X \subseteq X'$ and $Y \subseteq Y'$.*

6.4 Final steps

Here, we conclude the proof of the strong completeness, which is stated as Theorem 2 at the end of this section. The next three lemmas are auxiliary results used in the induction step of the proof of "induction" or "truth" Lemma 17. The most non-trivial of them is Lemma 15, whose proof is using the tumbled-pair construction.

Lemma 14 *For any epistemic world $w \in W$ and any formula $K_a\varphi \notin w$, there exists an epistemic world $u \in W$ such that $w \sim_a u$ and $\varphi \notin u$.*

Proof First, we show that the set of formulae $X = \{\neg\varphi\} \cup \{\psi \mid K_a\psi \in w\}$ is consistent. Assume the opposite, then there are formulae $K_a\psi_1, \ldots, K_a\psi_n \in w$ such that $\psi_1, \ldots, \psi_n \vdash \varphi$. Hence, $K_a\psi_1, \ldots, K_a\psi_n \vdash K_a\varphi$ by Lemma 6. Thus, $w \vdash K_a\varphi$ by the assumption $K_a\psi_1, \ldots, K_a\psi_n \in w$. Then, $K_a\varphi \in w$ because set w is maximal, which contradicts the assumption of the lemma. Therefore, set X is consistent.

By Lemma 3, there is a maximal consistent extension u of the set X. Note that $w \sim_a u$ by Definition 4 and the choice of sets X and u. Finally, $\neg\varphi \in X \subseteq u$ also by the choice of sets X and u. Therefore, $\varphi \notin u$ because set u is consistent. □

Lemma 15 *For any epistemic world $w \in W$ and any formula $\varphi \rhd_a \psi \notin w$, there exist epistemic worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $\varphi \in u$, $\psi \in u'$, and $u \not\sim_a u'$.*

Proof Consider the following two sets of formulae

$$X = \{\varphi\} \cup \{\chi \mid K_a\chi \in w\},$$
$$X' = \{\psi\} \cup \{\chi \mid K_a\chi \in w\}.$$

By Lemma 11 and the assumption $\varphi \rhd_a \psi \notin w$, pair (X, X') is (w, a)-tumbled. By Lemma 13, there is a (w, a)-saturated (w, a)-tumbled pair (Y, Y') such that $X \subseteq Y$ and $X' \subseteq Y'$. Sets Y and Y' are consistent by Lemma 9 and Lemma 10, respectively. Let sets u and u' be any maximal consistent extensions of sets Y and Y', respectively. Such sets exist by Lemma 3.

Note that by Definition 4 and the choice of sets X, Y, and u,

$$w \sim_a u. \tag{19}$$

Similarly, $w \sim_a u'$ by the choice of sets X', Y', and u'. Also, $\varphi \in X \subseteq Y \subseteq u$ and $\psi \in X' \subseteq Y' \subseteq u'$ by the choice of sets X, Y, u, X', Y', and u'.

To prove that $u \not\succ_a u'$, suppose the opposite. Thus, by Definition 5, there is a formula $\sigma \rhd_a \tau \in u$ such that

$$\sigma \in u \quad \text{and} \quad \tau \in u'. \tag{20}$$

The statement $\sigma \rhd_a \tau \in u$ implies that $u \vdash K_a(\sigma \rhd_a \tau)$ by the Introspection of Preference axiom and the Modus Ponens inference rule. Thus, $K_a(\sigma \rhd_a \tau) \in u$ because set u is maximal. Note, $u \sim_a w$ by Lemma 7 and statement (19). Hence, $\sigma \rhd_a \tau \in w$ by Definition 4. Recall that pair (Y, Y') is (w, a)-saturated by the choice of sets Y and Y'. Then, by Definition 8, either $\neg\sigma \in Y$ or $\neg\tau \in Y'$. Thus, either $\neg\sigma \in u$ or $\neg\tau \in u'$ because $Y \subseteq u$ and $Y' \subseteq u'$. Therefore, because sets u and u' are consistent, either $\sigma \notin u$ or $\tau \notin u'$, which contradicts statement (20). □

Lemma 16 *For any worlds $w, u, u' \in W$ and any formula $\varphi \rhd_a \psi \in w$, if $w \sim_a u$, $w \sim_a u'$, $\varphi \in u$, $\psi \in u'$, then $u \succ_a u'$.*

Proof The assumption $\varphi \rhd_a \psi \in w$ implies $w \vdash K_a(\varphi \rhd_a \psi)$ by the Introspection of Preference axiom and the Modus Ponens inference rule. Thus, $K_a(\varphi \rhd_a \psi) \in w$ because set w is maximal. Hence,

$$\varphi \rhd_a \psi \in u \tag{21}$$

by Definition 4 and the assumption $w \sim_a u$ of the lemma. At the same time, the assumptions $w \sim_a u$ and $w \sim_a u'$ imply that $u \sim_a u'$ by Lemma 7. Therefore, $u \succ_a u'$ by Definition 5, statement (21), and the assumptions $\varphi \in u$ and $\psi \in u'$ of the lemma. □

Lemma 17 *$w \Vdash \varphi$ iff $\varphi \in w$ for any world $w \in W$ and any formula $\varphi \in \Phi$.*

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

Suppose that formula φ has the form $K_a\psi$.

($\Rightarrow$) : Assume that $K_a\psi \notin w$. Thus, by Lemma 14, there exists an epistemic world $u \in W$ such that $w \sim_a u$ and $\psi \notin u$. Hence, $u \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash K_a\psi$ by item 4 of Definition 2.

($\Leftarrow$) : Let $K_a\psi \in w$. Consider any epistemic world $u \in W$ such that $w \sim_a u$. By item 4 of Definition 2, it suffices to show that $u \Vdash \psi$. Indeed, the assumptions $K_a\psi \in w$ and $w \sim_a u$ imply $\psi \in u$ by Definition 4. Therefore, $u \Vdash \psi$ by the induction hypothesis.

Finally, suppose that formula φ has the form $\psi \rhd_a \chi$.

($\Rightarrow$) : Assume that $\psi \rhd_a \chi \notin w$. Thus, by Lemma 15, there are epistemic worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $\psi \in u$, $\chi \in u'$, and $u \not\succ_a u'$. Then, $u \Vdash \psi$ and $u' \Vdash \chi$ by the induction hypothesis. Therefore, $w \nVdash \psi \rhd_a \chi$ by item 5 of Definition 2.

($\Leftarrow$) : Let $\psi \rhd_a \chi \in w$. Consider any epistemic worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \psi$, and $u' \Vdash \chi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$. Indeed, the assumptions $u \Vdash \psi$, and $u' \Vdash \chi$ imply $\psi \in u$ and $\chi \in u'$ by the induction hypothesis. Therefore, $u \succ_a u'$ by Lemma 16 and the assumptions $\psi \rhd_a \chi \in w$, $w \sim_a u$, and $w \sim_a u'$. □

We are now ready to state and prove the strong completeness theorem for our logical system.

Theorem 2 *For any set of formulae $X \subseteq \Phi$ and any formula $\varphi \in \Phi$, if $X \nvdash \varphi$, then there is a world w of an epistemic model with preferences such that $w \Vdash \chi$ for each formula $\chi \in X$ and $w \nVdash \varphi$.*

Proof The assumption $X \nvdash \varphi$ implies that the set $X \cup \{\neg\varphi\}$ is consistent. Thus, by Lemma 3, it can be extended to a maximal consistent set w. By Definition 3, set w is a world of the canonical epistemic model with preferences. Note that $\varphi \notin w$ because set w is consistent and $\neg\varphi \in X \subseteq w$. Therefore, $w \Vdash \chi$ for each formula $\chi \in X$ by Lemma 17. Also, $w \nVdash \varphi$ by the same Lemma 17. □

7 Discussion of the logical system

In the rest of the article, we discuss various properties of the proposed logical system and further compare our approach with the one existing in the literature.

7.1 Definability of knowledge through preferences

In this section, we prove that knowledge modality K_a is expressible through preference modality $\rhd_a$.

Theorem 3 $\vdash K_a\varphi \leftrightarrow ((\neg\varphi) \rhd_a \neg\varphi)$.

Proof We prove the two parts of the biconditional separately.
($\Rightarrow$) : Note that formula $\varphi \rightarrow (\neg\varphi \rightarrow \bot)$ is a propositional tautology. Hence, $\vdash K_a(\varphi \rightarrow (\neg\varphi \rightarrow \bot))$ by the Necessitation inference rule. Thus, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash K_a\varphi \rightarrow K_a(\neg\varphi \rightarrow \bot). \tag{22}$$

At the same time, the following formula is an instance of the first Monotonicity axiom:

$$\vdash K_a(\neg\varphi \rightarrow \bot) \rightarrow ((\bot \rhd_a \neg\varphi) \rightarrow ((\neg\varphi) \rhd_a \neg\varphi)).$$

Hence, by the laws of propositional reasoning using statement (22),

$$\vdash K_a\varphi \rightarrow ((\bot \rhd_a \neg\varphi) \rightarrow ((\neg\varphi) \rhd_a \neg\varphi)).$$

Then, again by the laws of propositional reasoning,

$$\vdash (\bot \rhd_a \neg\varphi) \to (K_a\varphi \to ((\neg\varphi) \rhd_a \neg\varphi)).$$

Therefore, $\vdash K_a\varphi \to ((\neg\varphi) \rhd_a \neg\varphi)$ by the Superiority of Falsehood axiom and the Modus Ponens inference rule.
($\Leftarrow$) : Observe that formula $((\neg\varphi) \rhd_a \neg\varphi) \to \neg(\neg\varphi \wedge \neg\varphi)$ is an instance of the Strictness axiom. Thus, $\vdash ((\neg\varphi) \rhd_a \neg\varphi) \to \varphi$ by the law of propositional reasoning. Hence, $\vdash K_a((\neg\varphi \rhd_a \neg\varphi) \to \varphi)$ by the Necessitation inference rule. Then, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash K_a((\neg\varphi) \rhd_a \neg\varphi) \to K_a\varphi. \tag{23}$$

At the same time, the following formula is an instance of the Introspection of Preference axiom:

$$((\neg\varphi) \rhd_a \neg\varphi) \to K_a((\neg\varphi) \rhd_a \neg\varphi).$$

Therefore, $\vdash ((\neg\varphi) \rhd_a \neg\varphi) \to K_a\varphi$ by the laws of propositional reasoning using statement (23). □

7.2 Preferences over non-exclusive statements

In this and the next sections, we compare our definition of preferences to the alternatives existing in the literature.

According to item 5 of Definition 2, an agent a has preferences for φ over ψ if, among all worlds indistinguishable from the current one, the agent prefers the worlds where φ is true to those where ψ is true. This means that statements φ and ψ must be *exclusive*. Indeed, if there is an indistinguishable world in which both statements are true, then the agent would prefer this world to itself. The letter is not possible because in Definition 1 we assume that the preference relation is strict.

To consider preferences between *non-exclusive* statements, one can alternatively require the agent to prefer the worlds where $\varphi \wedge \neg\psi$ is true to those where $\psi \wedge \neg\varphi$ is true. We denote this alternative preference modality by $\varphi \blacktriangleright_a \psi$. If the language Φ of our logical system is extended by the additional binary modality $\blacktriangleright_a$, then the definition of the satisfaction relation $\Vdash$ from Definition 2 should be extended as follows:

Definition 9 $w \Vdash \varphi \blacktriangleright_a \psi$, when for all epistemic worlds $u, u' \in W$, if $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi \wedge \neg\psi$, and $u' \Vdash \psi \wedge \neg\varphi$, then $u \succ_a u'$.

As discussed in the introduction, in the perfect information setting, such a definition of preferences has originally been proposed by Halldén (1957) under name "better". Doyle et al. (1991) call this notion "relative desire". Neither of them restrict worlds u and u' to those that are indistinguishable from the current world w because they consider settings with perfect information.

In this section, we illustrate the difference between our preference modality $\rhd$ and preference modality $\blacktriangleright$. We also compare expressive powers of these two modalities.

The next two observations illustrate the difference between the two preference modalities using the Battle of the Sexes with imperfect information example depicted in Fig. 3 (centre). To fit formulae in one line, in statements and proofs of both of these observations, by "one of them" we mean "**at least one of them**".

Observation 2 *For any epistemic world* $w \in \{oo, of, fo, ff\}$,

$$w \Vdash \textit{"One of them goes to opera"} \blacktriangleright_{\text{Brittany}} \textit{"One of them goes to football"}.$$

Proof Consider any worlds v, v' such that $w \sim_{\text{Brittany}} v$, $w \sim_{\text{Brittany}} v'$,

$$v \Vdash \text{"One of them goes to opera"} \wedge \neg\text{"One of them goes to football"}, \tag{24}$$
$$v' \Vdash \text{"One of them goes to football"} \wedge \neg\text{"One of them goes to opera"}. \tag{25}$$

By Definition 9, it suffices to show that $v \succ_{\text{Brittany}} v'$. Indeed, statement (24) and statement (25) imply that $v = oo$ and $v' = ff$, respectively. Therefore, $v \succ_{\text{Brittany}} v'$, see Fig. 3 (lower middle). □

Observation 3 *For any epistemic world* $w \in \{oo, of, fo, ff\}$,

$$w \nVdash \textit{"One of them goes to opera"} \rhd_{\text{Brittany}} \textit{"One of them goes to football"}.$$

Proof Note that $w \sim_{\text{Brittany}} of$ and $w \sim_{\text{Brittany}} ff$ because all worlds in our model are indistinguishable by Brittany, see Fig. 3 (lower middle). Also,

$$of \Vdash \text{"One of them goes to opera"},$$
$$ff \Vdash \text{"One of them goes to football"}$$

Finally, $of \not\succ_{\text{Brittany}} ff$, see Fig. 3 (lower middle). Therefore,

$$w \nVdash \text{"One of them goes to opera"} \rhd_{\text{Brittany}} \text{"One of them goes to football"}$$

by item 5 of Definition 2. □

Although one might argue whether modality $\rhd$ or modality $\blacktriangleright$ captures the notion of preference better, this question is not as important as it might seem. Indeed, as the next two theorems show, each of these modalities is expressible through the other one.

Theorem 4 $w \Vdash \varphi \blacktriangleright_a \psi$ *iff* $w \Vdash (\varphi \wedge \neg\psi) \rhd_a (\psi \wedge \neg\varphi)$ *for each epistemic world* w *of each epistemic model with preferences.*

Theorem 5 $w \Vdash \varphi \rhd_a \psi$ *iff* $w \Vdash K_a(\neg\varphi \vee \neg\psi) \wedge (\varphi \blacktriangleright_a \psi)$ *for each epistemic world* w *of each epistemic model with preferences.*

Proof ($\Rightarrow$) : Consider any world w of an arbitrary epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\succ_a\}_{a\in\mathcal{A}}, \pi)$. Suppose that $w \Vdash \varphi \rhd_a \psi$. It suffices to show that $w \Vdash K_a(\neg\varphi \vee \neg\psi)$ and $w \Vdash \varphi \blacktriangleright_a \psi$.

To prove that $w \Vdash K_a(\neg\varphi \vee \neg\psi)$, consider any epistemic world $u \in W$ such that $w \sim_a u$. By item 4 of Definition 2, it is enough to show that $u \Vdash \neg\varphi \vee \neg\psi$. Indeed, suppose the opposite. Then, $u \Vdash \varphi$ and $u \Vdash \psi$. Hence, $u \succ_a u$ by Definition 9, the assumption $w \sim_a u$, and the assumption $w \Vdash \varphi \rhd_a \psi$ of the theorem. Therefore, partial order $\succ_a$ is not strict, which contradicts item 3 of Definition 1.

To prove that $w \Vdash \varphi \blacktriangleright_a \psi$, consider any epistemic worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi \wedge \neg\psi$, and $u' \Vdash \psi \wedge \neg\varphi$. By Definition 9, it suffices to show that $u \succ_a u'$. Indeed, the statements $u \Vdash \varphi \wedge \neg\psi$ and $u' \Vdash \psi \wedge \neg\varphi$ imply that $u \Vdash \varphi$ and $u' \Vdash \psi$. Thus, $u \succ_a u'$ by the assumption $w \Vdash \varphi \rhd_a \psi$ of the theorem, item 5 of Definition 2, and the assumptions $w \sim_a u$ and $w \sim_a u'$.
($\Leftarrow$) : Suppose that $w \Vdash K_a(\neg\varphi \vee \neg\psi)$ and $w \Vdash \varphi \blacktriangleright_a \psi$. Towards the proof of $w \Vdash \varphi \rhd_a \psi$, consider any epistemic worlds $u, u' \in W$ such that

$$w \sim_a u, \tag{26}$$
$$w \sim_a u', \tag{27}$$
$$u \Vdash \varphi, \tag{28}$$
$$u' \Vdash \psi. \tag{29}$$

By Definition 9, it suffices to show that $u \succ_a u'$. Indeed, the assumption $w \Vdash K_a(\neg\varphi \vee \neg\psi)$, item 4 of Definition 2, and assumptions (26) and (27) imply that

$$u \Vdash \neg\varphi \vee \neg\psi, \tag{30}$$
$$u' \Vdash \neg\varphi \vee \neg\psi. \tag{31}$$

Statements (28) and (30) imply

$$u \Vdash \varphi \wedge \neg\psi. \tag{32}$$

Similarly, (29) and (31) imply

$$u' \Vdash \psi \wedge \neg\varphi. \tag{33}$$

Finally, by Definition 9, the assumption $w \Vdash \varphi \blacktriangleright_a \psi$ and statements (26), (27), (32), and (33) imply that $u \succ_a u'$. □

One might be concerned that Theorem 4 proves that modality $\blacktriangleright$ is definable through just modality $\rhd$, while Theorem 5 shows how to define modality $\rhd$ through modalities $\blacktriangleright$ and K. However, as the next theorem shows, modality K itself is definable through modality $\blacktriangleright$. Note that a similar result for modality $\rhd$ has already been established earlier as Theorem 3.

Theorem 6 *$w \Vdash K_a\varphi$ iff $w \Vdash (\varphi \blacktriangleright_a \neg\varphi) \wedge ((\neg\varphi) \blacktriangleright_a \varphi) \wedge \varphi$ for each epistemic world w of each epistemic model with preferences.*

Proof ($\Rightarrow$) : Suppose $w \nVdash (\varphi \blacktriangleright_a \neg\varphi) \wedge ((\neg\varphi) \blacktriangleright_a \varphi) \wedge \varphi$. Thus, one of the following cases takes place:
Case I: $w \nVdash \varphi \blacktriangleright_a \neg\varphi$. Thus, by Definition 9, there are epistemic worlds u, u' such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi \wedge \neg\neg\varphi$, $u' \Vdash \neg\varphi \wedge \neg\varphi$, and $u \not\succ_a u'$. Hence, $w \sim_a u'$ and $u' \Vdash \neg\varphi$. Therefore, $w \nVdash K_a\varphi$ by items 2 and 4 of Definition 2.
Case II: $w \nVdash (\neg\varphi) \blacktriangleright_a \varphi$. The proof in this case is similar to the proof in the case above.
Case III: $w \nVdash \varphi$. Therefore, $w \nVdash K_a\varphi$ by item 4 of Definition 2.
($\Leftarrow$) : Towards the contradiction, suppose that

$$w \Vdash \varphi \blacktriangleright_a \neg\varphi, \tag{34}$$

$$w \Vdash (\neg\varphi) \blacktriangleright_a \varphi, \tag{35}$$

$$w \Vdash \varphi, \tag{36}$$

$$w \nVdash K_a\varphi. \tag{37}$$

By item 4 of Definition 2, statement (37) implies that there is an epistemic world $u \in W$ such that $w \sim_a u$ and $u \nVdash \varphi$. Thus, taking into account statement (36),

$$w \Vdash \varphi \wedge \neg\neg\varphi \quad \text{and} \quad u \Vdash \neg\varphi \wedge \neg\varphi.$$

Hence, by Definition 9, statements (34) and (35) imply $w \succ_a u$ and $u \succ_a w$, respectively. Therefore, partial order $\succ_a$ is not strict, which contradicts item 3 of Definition 1. □

7.3 Betterness modality

In this article, we study the properties of the preferences expressible through the binary preference modality $\varphi \rhd_a \psi$. As discussed in the introduction, an alternative language for reasoning about preferences is proposed by Van Benthem et al. (2009) and it is later used in Christoff et al. (2021). Their language contains a unary "better" modality $[>]_a\varphi$. Informally, $[>]_a\varphi$ stands for "statement φ is true in all worlds that agent a prefers to the current world". Liu (2011, p.56) has combined this modality with the knowledge modality in a single-agent setting. Formally, the semantics of modality $[>]_a\varphi$ is captured in the following definition:

Definition 10 $w \Vdash [>]_a\varphi$, if $u \Vdash \varphi$ for each world u such that $u \succ_a w$.

In this section, we compare the expressive powers of the preference and the "better" modalities. Note that item 5 of Definition 2 specifies the semantics of $w \Vdash \varphi \rhd_a \psi$ in terms of the worlds indistinguishable from world w. At the same time, Definition 10 does *not* restrict the quantifier over u to the worlds indistinguishable from world w.

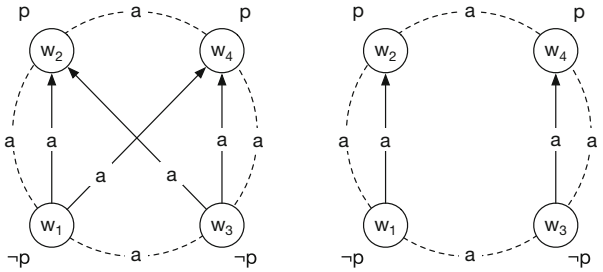

Fig. 4 Two Epistemic Models with Preferences

Thus, intuitively, modalities $\rhd_a$ and $[>]_a$ are not likely to be definable through each other. Perhaps, it is more interesting to compare the expressive powers of $\rhd$ and modality $[>]'_a$ that restricts u to the worlds indistinguishable from world w:

Definition 11 $w \Vdash [>]'_a\varphi$, if $u \Vdash \varphi$ for each world u such that $w \sim_a u$ and $u \succ_a w$.

We start with an observation by van Benthem et al. (2006) that modality $\rhd$ cannot be expressed through a combination of modalities $[>]$, $[>]'$, and K. Their original result is for perfect information models with non-strict partial order, but it could be easily adopted to our strict partial order setting with imperfect information. We state the result as Theorem 7 below. Our proof is a modified version of the original one from van Benthem et al. (2006). By Ψ we denote the language that contains modalities $[>]$, $[>]'$, and K, but does not contain modality $\rhd$. In other words, language Ψ is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid K_a \mid [>_a]\varphi \mid [>]'_a\varphi.$$

To prove Theorem 7, we construct two epistemic models with preferences indistinguishable in language Ψ, but distinguishable in language Φ.

The two models that we use are depicted in Fig. 4. We refer to these models as the left and the right ones. Without loss of generality, in this example, we assume that the set of propositional variables contains a single variable p and the set of agents $\mathcal{A}$ contains a single agent a. By $\succ^l$ and $\succ^r$ we denote the preference relations of the left and the right models, respectively. Similarly, by $\Vdash_l$ and $\Vdash_r$ we denote the satisfaction relations of these models.

We start the proof with an auxiliary observation.

Lemma 18 *For any formula $\varphi \in \Psi$,*

1. $w \Vdash_l [>]_a\varphi$ *iff* $w \Vdash_l [>]'_a\varphi$ *for any world* $w \in \{w_1, w_2, w_3, w_4\}$,
2. $w \Vdash_r [>]_a\varphi$ *iff* $w \Vdash_r [>]'_a\varphi$ *for any world* $w \in \{w_1, w_2, w_3, w_4\}$.

Proof Both parts of the lemma follow from Definition 10 and Definition 11 because all worlds in the left model and all worlds in the right model are indistinguishable by agent a. □

The next lemma shows that the left and the right models are indistinguishable in language Ψ. We state this lemma in a slightly more general form which is easier to prove by induction.

Lemma 19 *$w_i \Vdash_l \varphi$ iff $w_j \Vdash_r \varphi$ for any formula $\varphi \in \Psi$ and any integers $i, j \in \{1, 2, 3, 4\}$ such that $i \equiv j \pmod 2$.*

Proof We prove the statement of the lemma by structural induction on formula φ. If φ is a propositional variable, then the statement of the lemma follows from the definitions of the left and the right models, see Fig. 4. If formula φ is a negation or an implication, then the statement follows from the induction hypothesis and items 2 or 3 of Definition 2 in the standard way.

Assume that formula φ has the form $K_a\psi$.
($\Rightarrow$) : Suppose that $w_j \nVdash_r K_a\psi$. Thus, by item 4 of Definition 2, there is an integer $j' \in \{1, 2, 3, 4\}$ such that $w_{j'} \nVdash_r \psi$. Hence, $w_{j'} \nVdash_l \psi$ by the induction hypothesis. Therefore, $w_i \nVdash_l K_a\psi$ by item 4 of Definition 2. The proof in ($\Leftarrow$) direction is similar.

Assume that formula φ has the form $[>]_a\psi$. Recall that numbers i and j have the same parity by the assumption of the lemma. We consider the following two cases separately.
Case I: Numbers i and j are even. Then, there is no world u in the left model such that $u \succ_a^l w_i$, see Fig. 4 (left). Thus, $w_i \Vdash_l [>]_a\psi$ is vacuously true by Definition 10. Similarly, $w_j \Vdash_r [>]_a\psi$.
Case II: Numbers i and j are odd.
($\Rightarrow$) : Suppose that $w_j \nVdash_r [>]_a\psi$. Thus, by Definition 10, there is a world u in the right model such that $u \succ_a^r w_j$ and $u \nVdash_r \psi$. Then, the statement $u \succ_a^r w_j$ and the assumption of the case that j is odd imply that $u = w_{j+1}$, see Fig. 4 (right). Thus, $w_{j+1} \nVdash_r \psi$. Hence, $w_{j+1} \nVdash_l \psi$ by the induction hypothesis. Note that $w_{j+1} \succ_a^l w_j$ because number j is odd, see Fig. 4 (left). Therefore, $w_j \nVdash_l [>]_a\psi$ by Definition 10.
($\Leftarrow$) : Suppose that $w_i \nVdash_l [>]_a\psi$. Thus, by Definition 10, there is a world u in the left model such that $u \succ_a^l w_i$ and $u \nVdash_l \psi$. Note that statement $u \succ_a^l w_i$ implies that $u = w_k$ for some *even* integer k, see Fig. 4 (right). Hence, by the induction hypothesis,

$$w_2 \nVdash_r \psi \text{ and } w_4 \nVdash_r \psi. \tag{38}$$

Recall that integer j is odd by the assumption of the case. Then, $j + 1 \in \{2, 4\}$. Thus, $w_{j+1} \nVdash_r \psi$ by statement (38). Note that $w_{j+1} \succ_a^r w_j$ because j is odd, see Fig. 4 (right). Therefore, $w_j \nVdash_r [>]_a\psi$ by Definition 10.

Finally, if formula φ has the form $[>]'_a\psi$, then the statement of the lemma follows from the case $\varphi = [>]_a\psi$ by Lemma 18. □

The next two lemmas show that the left and the right models are distinguishable in language Φ.

Lemma 20 *$w \Vdash_l p \rhd_a \neg p$ for any world $w \in \{w_1, w_2, w_3, w_4\}$.*

Proof Consider any worlds $u, v \in \{w_1, w_2, w_3, w_4\}$ such that

$$u \Vdash_l p \quad \text{and} \quad v \Vdash_l \neg p. \tag{39}$$

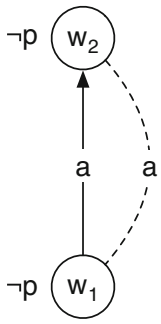

Fig. 5 An Epistemic Model with Preferences

By item 5 of Definition 2, it suffices to show that $u \succ_a^l v$. Indeed, assumptions (39) imply that $u \in \{w_2, w_4\}$ and $v \in \{w_1, w_3\}$, see Fig. 4. Therefore, $u \succ_a^l v$, see again Fig. 4. □

Lemma 21 *$w \nVdash_r p \rhd_a \neg p$ for any world $w \in \{w_1, w_2, w_3, w_4\}$.*

Proof Consider any world $w \in \{w_1, w_2, w_3, w_4\}$. Note that $w \sim_a^r w_2$, $w \sim_a^r w_3$, $w_2 \Vdash_r p$, $w_3 \Vdash_r \neg p$, and $w_2 \not\succ_a^r w_3$, see Fig. 4. Therefore, $w \nVdash_r p \rhd_a \neg p$ by item 5 of Definition 2. □

The next theorem follows from the three previous lemmas.

Theorem 7 *Modality $\rhd$ is not expressible in language Ψ over the class of all epistemic models with preferences.*

Next, we show that neither of the modalities $[>]$ and $[>]'$ is expressible in language Φ. To prove this, we consider the epistemic model with preferences depicted in Fig. 5. In Lemma 22, we show that worlds w_1 and w_2 of this model are not distinguishable in language Φ. In Lemma 23 and Lemma 24, we show that they are distinguishable using modalities $[>]$ and $[>]'$.

Lemma 22 *$w_1 \Vdash \varphi$ iff $w_2 \Vdash \varphi$ for any formula $\varphi \in \Phi$.*

Proof We prove the statement by induction on structural complexity of formula φ. If φ is a propositional variable p, then $w_1 \nVdash \varphi$ and $w_2 \nVdash \varphi$, see Fig. 5. The case when formula φ is a negation or an implication follows from the induction hypothesis and items 2 and 3 of Definition 2 in the standard way.

Suppose that formula φ has the form $K_a\psi$.

($\Rightarrow$) : Assume that $w_2 \nVdash K_a\psi$. Hence, by item 4 of Definition 2, there is a world u such that $w_2 \sim_a u$ and $u \nVdash \psi$. Note that $w_1 \sim_a w_2$, see Fig. 5. Hence, there is a world u such that $w_1 \sim_a u$ and $u \nVdash \psi$. Therefore, $w_1 \nVdash K_a\psi$ by item 4 of Definition 2. The proof of the case ($\Leftarrow$) is similar.[1]

Suppose that formula φ has the form $\psi_1 \rhd_a \psi_2$.

($\Rightarrow$) : Assume that $w_2 \nVdash \psi_1 \rhd_a \psi_2$. Hence, by item 5 of Definition 2, there are worlds u, v such that $w_2 \sim_a u$, $w_2 \sim_a v$, $u \Vdash \psi_1$, $v \Vdash \psi_2$, and $u \not\succ_a v$. Note that $w_1 \sim_a w_2$, see Fig. 5. Hence, there are worlds u, v such that $w_1 \sim_a u$, $w_1 \sim_a v$, $u \Vdash \psi_1$, $v \Vdash \psi_2$, and $u \not\succ_a v$. Therefore, $w_1 \nVdash \psi_1 \rhd_a \psi_2$. The proof of the case ($\Leftarrow$) is similar. □

[1] Alternatively, the case when φ has the form $K_a\psi$ can be shown using Theorem 3.

Lemma 23 *$w_1 \nVdash [>]p$ and $w_1 \nVdash [>]'p$.*

Proof Note that $w_2 \nVdash p$ and $w_2 \succ_a w_1$, see Fig. 5. Thus, $w_1 \nVdash [>]p$ by Definition 10. Note also that $w_1 \sim_a w_2$, see Fig. 5. Therefore, $w_1 \nVdash [>]'p$ by Definition 11. □

Lemma 24 *$w_2 \Vdash [>]p$ and $w_2 \Vdash [>]'p$.*

Proof To prove the first statement, note that there is no world u such that $u \succ_a w$, see Fig. 5. Thus, vacuously, $w_2 \Vdash [>]p$ by Definition 10. The proof of the second statement is similar. □

The next theorem follows from the three lemmas above.

Theorem 8 *Modalities $[>]_a$ and $[>]'_a$ are not expressible in language Φ over the class of all epistemic models with preferences.*

As we have seen in Theorem 7, modality $\rhd$ is not expressible in language Ψ over the class of **all** epistemic models with preferences. At the same time, as we show in the next theorem, $\rhd$ is expressible through modalities $[>]'$ and K over the class of epistemic models with **total** preference orders. For the setting where agents cannot distinguish any worlds (and, thus, K is the universal modality and modalities $[>]$ and $[>]'$ are equivalent), this result is stated in (Liu, 2011, p.39).

Theorem 9 *$w \Vdash \varphi \rhd_a \psi$ iff $w \Vdash K_a(\varphi \to \neg\psi \wedge [>]'_a\neg\psi)$ for any world w of any epistemic model with preferences such that strict order $\prec_a$ is **total**.*

Proof ($\Rightarrow$) : Consider any world u such that $w \sim_a u$. By item 4 of Definition 2, it suffices to show that $u \Vdash \varphi \to (\neg\psi \wedge [>]'_a\neg\psi)$. Suppose that $u \Vdash \varphi$. Then, by item 3 of Definition 2, it suffices to prove that $u \Vdash \neg\psi \wedge [>]'_a\neg\psi$. Suppose the opposite. Thus, by Definition 2, either $u \Vdash \psi$ or $u \nVdash [>]'_a\neg\psi$. We consider these two cases separately:
Case I: $u \Vdash \psi$. Then, $u \succ_a u$ by the assumption $w \Vdash \varphi \rhd_a \psi$ of the lemma, the assumptions $w \sim_a u$ and $u \Vdash \varphi$, and item 5 of Definition 2. Note that the statement $u \succ_a u$ contradicts $\succ_a$ being a strict order.
Case II: $u \nVdash [>]'_a\neg\psi$. Thus, by Definition 11, there is a world v such that $u \sim_a v$, $v \succ_a u$, and $v \nVdash \neg\psi$. Hence, $v \Vdash \psi$ by item 2 of Definition 2. Also, $w \sim_a v$ by the assumption $w \sim_a u$. By item 5 of Definition 2, the assumption $w \Vdash \varphi \rhd_a \psi$ of the lemma and the statements $w \sim_a u$, $w \sim_a v$, $u \Vdash \varphi$, and $v \Vdash \psi$ imply that $u \succ_a v$. The last statement contradicts the assumption $v \succ_a u$ because relation $\succ_a$ is a strict order.
($\Leftarrow$) : Consider any worlds $u, v \in W$ such that $w \sim_a u$, $w \sim_a v$, and

$$u \Vdash \varphi, \qquad v \Vdash \psi. \tag{40}$$

By item 5 of Definition 2, it suffices to show that $u \succ_a v$. Suppose the opposite:

$$u \not\succ_a v. \tag{41}$$

By item 4 of Definition 2, the assumption $w \Vdash K_a(\varphi \to \neg\psi \wedge [>]'_a\neg\psi)$ of the theorem and the assumption $w \sim_a u$ imply that

$$u \Vdash \varphi \to \neg\psi \wedge [>]'_a\neg\psi.$$

Thus, by the part $u \Vdash \varphi$ of statement (40) and Definition 2,

$$u \nVdash \psi, \qquad u \Vdash [>]'_a\neg\psi. \tag{42}$$

The part $v \Vdash \psi$ of statement (40) and the part $u \nVdash \psi$ of statement (42) imply that $u \neq v$. Hence, $v \succ_a u$ by statement (41) and the assumption of the theorem that order $\succ_a$ is total. Then, the part $u \Vdash [>]'_a\neg\psi$ of statement (42) implies that $v \Vdash \neg\psi$ by Definition 11 and the assumptions $w \sim_a u$ and $w \sim_a v$. Therefore, $v \nVdash \psi$ by item 2 of Definition 2, which contradicts the part $v \Vdash \psi$ of statement (40). □

As we have seen in the previous theorem, over the class of all epistemic models with **total** preference order, modality $\rhd$ is expressible through modalities $[>]'$ and K. We conclude this section with the observation that the opposite is not true.

Theorem 10 *Modalities $[>]_a$ and $[>]'_a$ are not expressible in language Φ over the class of epistemic models with preferences in which strict order $\succ_a$ is **total**.*

Proof The proof of this theorem is identical to the proof of Theorem 8 because order $\succ_a$ in Fig. 5 is total. □

7.4 Non-strict preferences

The preferences that we consider in this article are strict. There are at least four ways in which non-strictness could be introduced into our setting.

First Way Item 3 of Definition 1 requires relation to be a strict partial order. In other words, this item requires the relation to be irreflexive and transitive. If the irreflexivity condition is removed, then the Strictness axiom is no longer sound. Note, however, that this axiom is used only once in the proof of completeness. Namely, it is applied in Lemma 8 to show irreflexivity of relation $\succ_a$ in the canonical model. Thus, if the irreflexivity condition and the Strictness axiom are removed, then the existing proof of completeness remains valid. Hence, our logical system (without the Strictness axiom) is sound and complete with respect to the modified semantics.

It is interesting to point out that the Strictness Axiom is also used in the proof of Theorem 3 to establish definability of knowledge through preferences. It is easy to see that if the Strictness axiom and the irreflexivity assumption of Item 3 of Definition 1 are omitted, then not only the proof of Theorem 3 is not valid, but even the statement of the theorem is not true. We think that, in this case, knowledge is not definable through preferences, but we have not tried to prove this.

Second Way The requirement of irreflexivity in item 3 of Definition 1 can be replaced with reflexivity and antisymmetry. This would mean that relation $\succ$ is a partial order. In this case, the Strictness axiom will still not be valid, but the existing proof of

completeness cannot be easily modified to work in the new setting. The problem comes from the fact that we don't know how to make relation $\succ$ of the canonical model to be antisymmetric. As a result, we cannot prove a completeness theorem in this case.

In this setting, the statement of Theorem 3 remains false, but the modality $K_a\varphi$ is definable via preferences as $(\varphi \rhd_a (\neg\varphi)) \wedge ((\neg\varphi) \rhd_a \varphi) \wedge \varphi$. Note that this way to define knowledge through preferences is not valid in the setting of above First Way.

Third Way Another option is to keep Definition 1 unchanged. Instead, one can first define relation $\succeq_a$ as a reflexive closure of relation $\succ_a$. In other words, $w \succeq_a u$ is true if either $w = u$ or $w \succeq_a u$. Then, item 5 of Definition 2 can be adjusted to use relation $\succeq_a$ instead of relation $\succ_a$. It is easy to see that the logical system created by this change is the same as in Second Way. All that we said above applies here as well.

Fourth Way Finally, without any changes to the existing definitions, a non-strict preference modality $\varphi \trianglerighteq_a \psi$ can be defined as $K_a(\varphi \leftrightarrow \psi) \vee (\varphi \rhd_a \psi)$. It is interesting to note that the original strict preference modality $\varphi \rhd_a \psi$ is definable via K_a and $\trianglerighteq_a$ as $\neg K_a(\varphi \leftrightarrow \psi) \wedge \varphi \trianglerighteq_a \psi$. Because of this mutual definability, expressive power of the language containing modalities K_a and $\trianglerighteq_a$ is the same as of our original language. Thus, most of the results of this article, including the completeness theorem, apply to the language with modalities K_a and $\trianglerighteq_a$. To conclude, note that $K_a\varphi$ is equivalent in this setting to $\varphi \trianglerighteq_a \top$, where, as usual, $\top$ is $\neg\bot$.

7.5 Preferences and public announcements

Public Announcement Logic (PAL) is a popular example of a dynamic epistemic logic (van Ditmarsch et al., 2007; Pacuit, 2013) that combines knowledge modality K_a with public announcement modality $[\chi]\varphi$. Informally, the expression $[\chi]\varphi$ stands for "statement φ is true after a public announcement of truthful statement χ". Multiple extensions of Public Announcement Logic are suggested. Wáng and Ågotnes (2013) add to it the distributed knowledge modality. Ågotnes et al. (2010) propose a group announcement modality $\langle G\rangle\varphi$ that means "group G can announce certain facts, individually known to members of the group, after which statement φ will be true". Galimullin and Alechina (2017) study a modality $\langle G\rangle\varphi$ that states "φ will become true after an announcement by group G and will remain true after further announcements made by agents outside of group G".

The logic of epistemic preferences can be extended with the public announcement modality $[\varphi]\psi$. The semantics of such an extension combines Definition 2 with the standard semantics of PAL (van Ditmarsch et al., 2007). It is a well-known observation that after a public announcement of a true statement χ, this statement might no longer be true. It is interesting to note that a public announcement might not only affect agent's preferences, but it can even change them to the opposite. Namely, below we construct an epistemic model with preferences and three statements φ, ψ, and χ such that $\varphi \rhd_a \psi$ and $[\chi](\psi \rhd_a \varphi)$ in some world of the model.

The model that we consider captures the situation in which an agent a tosses a coin while being fearful of a world apocalypse happening tomorrow. The agent is indifferent to the outcome of the coin toss, but he would prefer the apocalypse not

to happen. Formally, our model contains the following four worlds indistinguishable by the agent: $(apocalypse, heads)$, $(apocalypse, tails)$, $(no\,apocalypse, heads)$, $(no\,apocalypse, tails)$. Consider the following three statements:

$$\varphi \equiv K_a(\text{“It's heads”}) \leftrightarrow \text{“World apocalypse is tomorrow”}, \tag{43}$$

$$\psi \equiv K_a(\text{“It's heads”}) \leftrightarrow \neg\text{“World apocalypse is tomorrow”},$$

$$\chi \equiv \text{“It's heads”}. \tag{44}$$

Suppose that the coin lands heads up. Note that before the truthful announcement χ, agent a does not know the result of the toss. Thus, the statement K_a(“It's heads”) is false. Hence, due to statements (43) and (44), *before the announcement*, agent a knows that φ is equivalent to

$$\neg\text{“World apocalypse is tomorrow”}$$

while ψ is equivalent to “World apocalypse is tomorrow”. Then, $\varphi \rhd_a \psi$ because the agent would prefer the apocalypse not to happen.

At the same time, after the truthful public announcement of χ, the statement K_a(“It's heads”) is true. Hence, again due to statements (43) and (44), *after the announcement*, agent a knows that φ is equivalent to “World apocalypse is tomorrow” while ψ is equivalent to $\neg$“World apocalypse is tomorrow”. Thus, $[\chi](\psi \rhd_a \varphi)$ again because the agent would prefer the apocalypse not to happen. Therefore, public announcement of χ changes the preference of the agent to the opposite.

8 Conclusion

In this article, we have proposed a new approach to defining preference in the imperfect information setting. We say that an agent prefers φ over ψ if, among all indistinguishable worlds, he prefers those where φ is true to those where ψ is true. We have captured this definition as a binary modality and compared our approach to several others existing in the literature. Our main technical result is a complete logical system describing the interplay between the preference modality and the individual knowledge modality in the multiagent setting. The proof of completeness theorem is using a newly proposed “tumbled pairs” construction.

Declarations

Conflict of Interests The authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Proof of Soundness

The soundness of the Truth, the Negative Introspection, and the Distributivity axioms are well-known. We prove the soundness of each of the remaining axioms as a separate lemma.

Lemma 25 *If* $w \Vdash \varphi \rhd_a \psi$, *then* $w \Vdash K_a(\varphi \rhd_a \psi)$.

Proof Consider any world $w' \in W$ such that $w \sim_a w'$. By item 4 of Definition 2, it suffices to show that $w' \Vdash \varphi \rhd_a \psi$. Towards this proof, consider any worlds $u, u' \in W$ such that $w' \sim_a u$, $w' \sim_a u'$, $u \Vdash \varphi$, and $u' \Vdash \psi$. By item 5 of Definition 2, it suffices to prove that $u \succ_a u'$.

Indeed, the assumptions $w \sim_a w'$, $w' \sim_a u$, and $w' \sim_a u'$ imply that $w \sim_a u$ and $w \sim_a u'$. Thus, $u \succ_a u'$ by item 5 of Definition 2, the assumption $w \Vdash \varphi \rhd_a \psi$ of the lemma, and the assumptions $u \Vdash \varphi$ and $u' \Vdash \psi$. □

Lemma 26 *If* $w \Vdash K_a(\varphi \to \psi)$ *and* $w \Vdash \psi \rhd_a \chi$, *then* $w \Vdash \varphi \rhd_a \chi$.

Proof Consider any worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi$, and $u' \Vdash \chi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$.

The assumption $w \sim_a u$ implies $u \Vdash \varphi \to \psi$ by item 4 of Definition 2 and the assumption $w \Vdash K_a(\varphi \to \psi)$ of the lemma. Then, $u \Vdash \psi$ by the assumption $u \Vdash \varphi$ and item 3 of Definition 2. Therefore, $u \succ_a u'$ by the assumption $w \Vdash \psi \rhd_a \chi$ of the lemma and the assumptions $w \sim_a u$, $w \sim_a u'$, and $u' \Vdash \chi$. □

Lemma 27 *If* $w \Vdash K_a(\varphi \to \psi)$ *and* $w \Vdash \chi \rhd_a \psi$, *then* $w \Vdash \chi \rhd_a \varphi$.

Proof Consider any worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \chi$, and $u' \Vdash \varphi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$.

The assumption $w \sim_a u'$ implies $u' \Vdash \varphi \to \psi$ by item 4 of Definition 2 and the assumption $w \Vdash K_a(\varphi \to \psi)$ of the lemma. Then, $u' \Vdash \psi$ by the assumption $u' \Vdash \varphi$ and item 3 of Definition 2. Therefore, $u \succ_a u'$ by the assumption $w \Vdash \chi \rhd_a \varphi$ of the lemma and the assumptions $w \sim_a u$, $w \sim_a u'$, and $u \Vdash \chi$. □

Lemma 28 *If* $w \Vdash \varphi \rhd_a \psi$, *then* $w \nVdash \varphi \wedge \psi$.

Proof Suppose that $w \Vdash \varphi \wedge \psi$. Thus, $w \Vdash \varphi$ and $w \Vdash \psi$. Hence, $w \succ_a w$ by item 5 of Definition 2 and the assumption $w \Vdash \varphi \rhd_a \psi$ of the lemma. Therefore, relation $\succ_a$ is not a strict partial order, which contradicts item 3 of Definition 1. □

Lemma 29 $w \Vdash \bot \rhd_a \varphi$.

Proof Consider any worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \bot$, and $u' \Vdash \varphi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$. This is vacuously true because there is no world $u \in W$ such that $u \Vdash \bot$. □

The proof of the next lemma is similar to the previous one.

Lemma 30 $w \Vdash \varphi \rhd_a \bot$.

Lemma 31 *If* $w \Vdash \varphi \rhd_a \psi$, $w \Vdash \psi$, *and* $w \Vdash \psi \rhd_a \chi$, *then* $w \Vdash \varphi \rhd_a \chi$.

Proof Consider any worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi$, and $u' \Vdash \chi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$.

Note that $w \sim_a w$. Thus, $u \succ_a w$ by item 5 of Definition 2, the assumption $w \Vdash \varphi \rhd_a \psi$ of the lemma, the assumptions $w \sim_a u$ and $u \Vdash \varphi$, and the assumption $w \Vdash \psi$ of the lemma.

Similarly, $w \succ_a u'$ by item 5 of Definition 2, the assumptions $w \Vdash \psi \rhd_a \chi$ and $w \Vdash \psi$ of the lemma, and the assumptions $w \sim_a u'$ and $u' \Vdash \chi$.

Finally, note that the statements $u \succ_a w$ and $w \succ_a u'$ imply that $u \succ_a u'$ because relation $\succ_a$ is transitive. □

Lemma 32 *If* $w \Vdash \varphi \rhd_a \chi$ *and* $w \Vdash \psi \rhd_a \chi$, *then* $w \Vdash (\varphi \vee \psi) \rhd_a \chi$.

Proof Consider any worlds $u, u' \in W$ such that $w \sim_a u$, $w \sim_a u'$, $u \Vdash \varphi \vee \psi$, and $u' \Vdash \chi$. By item 5 of Definition 2, it suffices to show that $u \succ_a u'$. Indeed, the statement $u \Vdash \varphi \vee \psi$ implies that either $u \Vdash \varphi$ or $w \Vdash \psi$. If $u \Vdash \varphi$, then $u \succ_a u'$ by the assumption $w \Vdash \varphi \rhd_a \chi$ of the lemma, item 5 of Definition 2, and the assumptions $w \sim_a u$, $w \sim_a u'$, and $u' \Vdash \chi$. The case $w \Vdash \psi$ is similar. □

The proof of the next lemma is similar to the previous one.

Lemma 33 *If* $w \Vdash \varphi \rhd_a \psi$ *and* $w \Vdash \varphi \rhd_a \chi$, *then* $w \Vdash \varphi \rhd_a (\psi \vee \chi)$.

Auxiliary Lemma

Lemma 7 *Relation* $\sim_a$ *is an equivalence relation on set* W *for each* $a \in \mathcal{A}$.

Proof **Reflexivity:** Consider any formula $\varphi \in \Phi$. Suppose that $K_a\varphi \in w$. It suffices to show that $\varphi \in w$. Indeed, the assumption $K_a\varphi \in w$ implies $w \vdash \varphi$ by the Truth axiom and the Modus Ponens inference rule. Therefore, $\varphi \in w$ because set w is maximal.
Symmetry: Consider any epistemic worlds $w, u \in W$ such that $w \sim_a u$ and any formula $K_a\varphi \in u$. It suffices to show that $\varphi \in w$. Suppose the opposite. Then, $\varphi \notin w$. Hence, $w \nvdash \varphi$ because set w is maximal. Thus, $w \nvdash K_a\varphi$ by the contraposition of the Truth axiom. Then, $\neg K_a\varphi \in w$ because set w is maximal. Thus, $w \vdash K_a\neg K_a\varphi$ by the Negative Introspection axiom and the Modus Ponens inference rule. Hence, $K_a\neg K_a\varphi \in w$ because set w is maximal. Then, $\neg K_a\varphi \in u$ by the assumption $w \sim_a u$ and Definition 4. Thus, $K_a\varphi \notin u$ because set u is consistent, which contradicts the assumption $K_a\varphi \in u$.
Transitivity: Consider any worlds $w, u, v \in W$ such that $w \sim_a u$ and $u \sim_a v$ and any formula $K_a\varphi \in w$. It suffices to show that $\varphi \in v$. The assumption $K_a\varphi \in w$ implies $w \vdash K_aK_a\varphi$ by Lemma 4 and the Modus Ponens inference rule. Thus, $K_aK_a\varphi \in w$ because set w is maximal. Hence, $K_a\varphi \in u$ by the assumption $w \sim_a u$ and Definition 4. Therefore, $\varphi \in v$ by the assumption $u \sim_a v$ and Definition 4. □

References

Ågotnes, T., Balbiani, P., van Ditmarsch, H., & Seban, P. (2010). Group announcement logic. *Journal of Applied Logic, 8*(1), 62–81. https://doi.org/10.1016/j.jal.2008.12.002

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

Bruckner, D. W. (2009). In defense of adaptive preferences. *Philosophical Studies, 142*(3), 307–324.

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

Doyle, J., Shoham, Y., & Wellman, M. P. (1991). A logic of relative desire. In *International Symposium on Methodologies for Intelligent Systems* (pp. 16–31). Springer

Galimullin, R., & Alechina, N. (2017). Coalition and group announcement logic. In *Proceedings Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK) 2017*, Liverpool, UK, 24–26 July 2017, pp. 207–220

Grossi, D., van der Hoek, W., & Kuijer, L. B. (2022). Reasoning about general preference relations. *Artificial Intelligence, 313*, 103793.

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

Jiang, J., & Naumov, P. (2022). The egocentric logic of preferences. In *The 31st International Joint Conference on Artificial Intelligence (IJCAI-22)*

Liu, F. (2011). *Reasoning about preference dynamics* (Vol. 354). Springer Science & Business Media.

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

Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT Press.

Pacuit, E. (2013). Dynamic epistemic logic II: Logics of information change. *Philosophy Compass, 8*(9), 815–833.

van Benthem, J., van Otterloo, S., & Roy, O. (2006). Preference logic, conditionals and solution concepts in games. In: H. Lagerlund, S. Lindström, R. Sliwinski (eds.) Modality matters: twenty-five essays in honour of Krister Segerberg, pp. 61–77. Uppsala Univ., Dept. of Philosophy (Uppsala Philosophical Studies 53)

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.

van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*. Springer. https://doi.org/10.1007/978-1-4020-5839-4

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

Wáng, Y. N., & Ågotnes, T. (2013). Public announcement logic with distributed knowledge: Expressivity, completeness and complexity. *Synthese, 190*(1), 135–162.