**L**ogic and **L**ogical **P**hilosophy
Online First Articles (2024)
DOI: 10.12775/LLP.2024.009

**Sanaz Azimipour**
**Pavel Naumov**

# If You're Happy, Then You Know It: The Logic of Happiness … and Sadness

**Abstract.** The article proposes a formal semantics of happiness and sadness modalities in the imperfect information setting. It shows that these modalities are not definable through each other and gives a sound and complete axiomatization of their properties.

**Keywords**: epistemic logic; preferences; completeness; undefinability

## 1. Introduction

Different formal models of human emotions have been studied in the literature. Doyle, Shoham, and Wellman propose a logic of relative desire [6]. Lang, van der Torre, and Weydert introduce utilitarian desires [9]. Meyer states logical principles aiming at capturing anger and fear [13]. Steunebrink, Dastani, and Meyer expand this work to hope [16]. Adam, Herzig, and Longin propose formal definitions of hope, fear, relief, disappointment, resentment, gloating, pride, shame, admiration, reproach, gratification, remorse, gratitude, and anger [1]. Lorini and Schwarzentruber define regret and elation [11].

The focus of this article is on happiness and sadness. These notions have long been studied in literature on philosophy [2, 4, 7, 18], psychology [3, 17], and economics [5, 8]. Note that happiness/sadness are vague terms that have multiple meanings that overlap with several other terms such as joy/distress and elation/disappointment.

Two approaches to capturing happiness and sadness in formal logical systems have been proposed. The first approach is based on Ortony,

*Received* July 6, 2023. *Accepted* December 20, 2023. *Published online* February 5, 2024

Clore, and Collins' definitions of joy and distress (the capitalization is original):

> [...] we shall often use the terms "joy" and "distress" as convenient shorthands for the reactions of being PLEASED ABOUT A DESIRABLE EVENT and DISPLEASED ABOUT AN UNDESIRABLE EVENT, respectively. [14, p. 88]

Adam, Herzig, and Longin formalized these definitions. An agent feels joy about $\varphi$ if she believes that $\varphi$ is true and she desires $\varphi$. An agent feels distress about $\varphi$ if she believes $\varphi$ is true, but she desires $\varphi$ to be false [1]. Similarly, Lorini and Schwarzentruber define that an agent is elated/disappointed about $\varphi$ if $\varphi$ is desirable/undesirable to the agent, the agent knows that $\varphi$ is true, and she also knows that the others could have prevented $\varphi$ from being true [11]. Although Adam, Herzig, and Longin use beliefs while Lorini and Schwarzentruber use knowledge and the latter authors also add "could have prevented" part, both definitions could be viewed as a variation of Ortony, Clore, and Collins' definitions of joy/distress.

Meyer suggested a very different approach to defining these notions. He writes "an agent that is happy observes that its subgoals (towards certain goals) are being achieved, and is 'happy' with it". He acknowledges, however, that this definition might be capturing only one of the forms of what people mean by happiness [13].

Note, for example, that if Pavel, the second author of this article, receives an unexpected gift from Sanaz, the first author, then he will experience "joy" as defined by Ortony, Clore, and Collins. However, he will not be "happy" as defined by Meyer because receiving such a gift has never been among Pavel's goals[1].

In this article, we adopt Ortony, Clore, and Collins' definitions, but we use the terms happiness/sadness instead of joy/distress. While the cited above works suggest formal semantics and list formal properties of happiness and sadness, none of them gives an axiomatization of these properties. In this article, we propose such an axiomatization and prove its completeness. We also show that notions of happiness and sadness in our formal system are, in some sense, dual but are not definable through each other.

[1] Ortony, Clore, and Collins give a similar example with an unexpected inheritance from an unknown relative.

The rest of the article is structured as follows. First, we formally define epistemic models with preferences that serve as the foundation of our semantics of happiness and sadness. Then, we define the formal syntax and semantics of our system and illustrate them with several examples. In Section 7, we show that sadness cannot be defined through happiness. In spite of this, as we show in Section 8 there is a certain duality between the properties of happiness and sadness. We use this duality to observe that happiness can not be defined through sadness either. In Section 9, we list the axioms of our logical system. In the section that follows, we prove its soundness. In Sections 11 and 12, we show how utilitarian and goodness-based approaches to desires, already existing in the literature, could be adapted to happiness and sadness. In the rest of the article, we prove the completeness of our logical section. The last section provides a conclusion.

## 2. Epistemic Models with Preferences

Throughout the article, we assume a fixed countable set of agents $\mathcal{A}$ and a countable set of propositional variables. The semantics of our logical system is defined in terms of epistemic models with preferences. These models extend standard Kripke models for epistemic logic with a preference relation for each agent in set $\mathcal{A}$.

Definition 2.1. A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$ is called 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. $\prec_a$ is a strict partial order preference relation on set $W$ for each agent $a \in \mathcal{A}$,
4. $\pi(p)$ is a subset of $W$ for each propositional variable $p$.

We read $w \prec_a u$ as "an agent $a$ prefers a world $u$ over a world $w$". For any two sets of epistemic worlds $U, V \subseteq W$, we write $U \prec_a V$ if $u \prec_a v$ for each world $u \in U$ and each world $v \in V$.

An example of an epistemic model with preferences is depicted in Figure 1. It captures the mentioned in the introduction scenario in which the first author of this article, Sanaz, is sending a gift to the second author, Pavel. If Pavel receives the gift, he will acknowledge it by sending

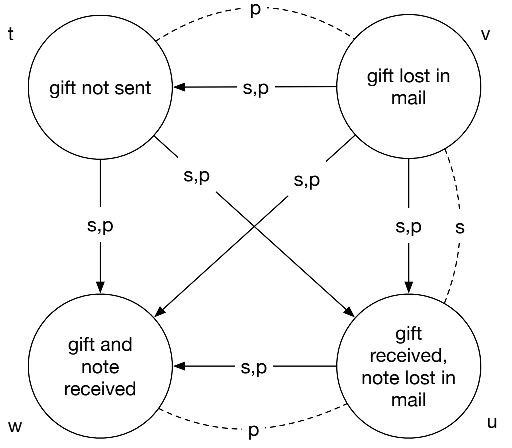


Figure 1. Gift Scenario.

back a thank-you card. We assume that either the gift or the card can be lost in the mail and that there is no additional communication between the authors.

This epistemic model with preferences in Figure 1 has four worlds corresponding to four different scenarios. In the world $t$ Sanaz did not send the gift. In the world $v$, she sent the gift, but it was lost in the mail. In the world $u$, Pavel received the gift, but his card is lost in the mail. Finally, in the world $w$ Pavel received the gift and Sanaz received his card. Sanaz cannot distinguish the world $v$, in which the gift is lost, from the world $u$, in which the card is lost. Pavel cannot distinguish the world $t$, in which the gift is not sent, from the world $v$, in which the gift is lost. He also cannot distinguish the world $u$, in which the card is lost, from the world $w$, in which the card is received. In the diagram, the indistinguishability relations of Sanaz and Pavel are denoted by dashed lines labelled by $s$ and $p$ respectively. Although in general, according to Definition 2.1, different agents might have different preference relations between worlds, in this scenario we assume that Sanaz and Pavel have the same preferences. These preferences are shown in the diagram using directed edges. For example, the directed edge from $v$ to $t$ labelled with $s, p$ means that they both would prefer if Sanaz does not send the gift at all to the scenario when the gift is lost in the mail.

The preference relation could be used to capture the agent's utility function, desires, motivations, goals, and intentions. For instance, in the example depicted in Figure 1, Sanaz's preferences capture her intentions for Pavel to receive the gift when she sends it in the mail.

## 3. Syntax and Semantics

In this section, we introduce the formal syntax and semantics of our logical system. Throughout the article, we assume a fixed countable set of agents $\mathcal{A}$ and a fixed nonempty countable set of propositional variables. The language $\Phi$ of our logical system is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{N}\varphi \mid \mathsf{K}_a\varphi \mid \mathsf{H}_a\varphi \mid \mathsf{S}_a\varphi,$$

where $p$ is a propositional variable and $a \in \mathcal{A}$ is an agent. We read the formula $\mathsf{N}\varphi$ as "a statement $\varphi$ is true in every world", the formula $\mathsf{K}_a\varphi$ as "an agent $a$ knows $\varphi$", the formula $\mathsf{H}_a\varphi$ as "an agent $a$ is happy about $\varphi$", and the formula $\mathsf{S}_a\varphi$ as "an agent $a$ is sad about $\varphi$". We assume that Boolean connectives conjunction $\wedge$, disjunction $\vee$, and biconditional $\leftrightarrow$ are defined through negation $\neg$ and implication $\to$ in the standard way. By $\overline{\mathsf{N}}\varphi$ we denote the formula $\neg\mathsf{N}\neg\varphi$. We read $\overline{\mathsf{N}}\varphi$ as "a statement $\varphi$ is true in at least one of the worlds".

Definition 3.1. For any world $w \in W$ of an epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$ and any formula $\varphi \in \Phi$, satisfaction relation $w \Vdash \varphi$ is defined 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. $w \Vdash \mathsf{N}\varphi$, if $u \Vdash \varphi$ for each world $u \in W$,
5. $w \Vdash \mathsf{K}_a\varphi$, if $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
6. $w \Vdash \mathsf{H}_a\varphi$, if the following three conditions are satisfied:
   (a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
   (b) for any two worlds $u, u' \in W$, if $u \nVdash \varphi$ and $u' \Vdash \varphi$, then $u \prec_a u'$,
   (c) there is a world $u \in W$ such that $u \nVdash \varphi$,
7. $w \Vdash \mathsf{S}_a\varphi$, if the following three conditions are satisfied:
   (a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
   (b) for any two worlds $u, u' \in W$, if $u \Vdash \varphi$ and $u' \nVdash \varphi$, then $u \prec_a u'$,
   (c) there is a world $u \in W$ such that $u \nVdash \varphi$.

Items 6 and 7 of Definition 3.1 capture the notions of happiness and sadness studied in this article. Item 6 states that to be happy about a condition $\varphi$, the agent must know that $\varphi$ is true, the agent must prefer the worlds in which condition $\varphi$ is true to those where it is false, and the condition $\varphi$ must not be trivial. These three parts are captured by items 6(a), 6(b), and 6(c) of the above definition. Note that we require condition $\varphi$ to be non-trivial to exclude an agent from being happy about conditions that always hold in the model. Thus, for example, we believe that an agent cannot be happy that $2+2=4$.

Similarly, item 7 states that an agent is sad about condition $\varphi$ if she knows that $\varphi$ is true, she prefers worlds in which condition $\varphi$ is false to those in which condition $\varphi$ is true, and condition $\varphi$ is not trivial. Note that being sad is different from not being happy. In fact, later in this article we show that neither of modalities $\mathsf{H}$ and $\mathsf{S}$ is expressible through the other.

We conclude this section with a technical observation that follows from Definition 3.1. Informally, it states that if two formulae are satisfied in the same worlds of a model, then these formulae evoke the same emotions in all worlds of the model.

Lemma 3.1. *For any agent $a \in \mathcal{A}$, any formulae $\varphi, \psi \in \Phi$, and any epistemic model with preferences $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$, if $w \Vdash \varphi$ iff $w \Vdash \psi$ for each world $w \in W$, then'*

1. *$w \Vdash \mathsf{H}_a\varphi$ iff $w \Vdash \mathsf{H}_a\psi$ for each world $w \in W$,*
2. *$w \Vdash \mathsf{S}_a\varphi$ iff $w \Vdash \mathsf{S}_a\psi$ for each world $w \in W$.* $\dashv$

## 4. The Gift Scenario

In this section, we illustrate Definition 3.1 using the gift scenario depicted in the diagram in Figure 1. For the convenience of the reader, we reproduce this diagram in Figure 2.

Proposition 4.1. *$z \Vdash \mathsf{H}_p$("Pavel received a gift from Sanaz") if and only if $z \in \{w, u\}$.*

Proof. ($\Rightarrow$) Suppose $z \notin \{w, u\}$. Thus, $z \in \{t, v\}$. Hence, see Figure 2,

$$z \nVdash \text{"Pavel received a gift from Sanaz"}.$$

Therefore, $z \nVdash \mathsf{H}_p$("Pavel received a gift from Sanaz") by item 6(a) of Definition 3.1 because $z \sim_p z$.

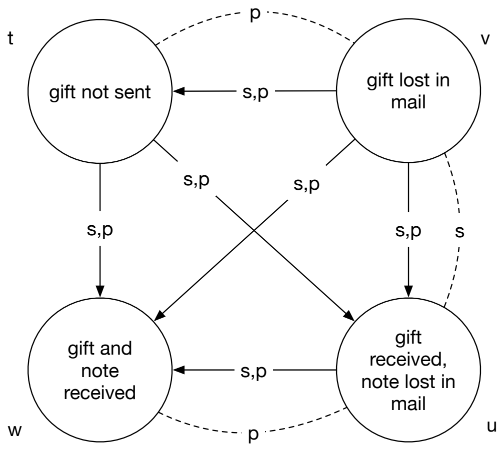


Figure 2. Gift Scenario (repeated from Figure 1).

($\Leftarrow$) Let $z \in \{w, u\}$. To show

$$z \Vdash \mathsf{H}_p(\text{“Pavel received a gift from Sanaz”}),$$

we will verify conditions (a), (b) and (c) of item 6 of Definition 3.1 separately:

*Condition a*: Consider any world $z'$ such that $z \sim_p z'$. To verify the condition, it suffices to show that $z' \Vdash$ “Pavel received a gift from Sanaz”. Indeed, assumptions $z \in \{w, u\}$ and $z \sim_p z'$ imply that $z' \in \{w, u\}$, see Figure 2. Therefore, again see Figure 2,

$$z' \Vdash \text{“Pavel received a gift from Sanaz”}.$$

*Condition b*: Consider any two epistemic worlds $x, y$ such that

$$x \nVdash \text{“Pavel received a gift from Sanaz”}, \tag{4.1}$$

$$y \Vdash \text{“Pavel received a gift from Sanaz”}. \tag{4.2}$$

To verify the condition, it suffices to show that $x \prec_p y$. Indeed, assumptions (4.1) and (4.2) implies that $x \in \{t, v\}$ and $y \in \{w, u\}$, see Figure 2. Note that $\{t, v\} \prec_p \{w, u\}$, see Figure 2. Therefore, $x \prec_p y$.

*Condition c*: $t \nVdash$ “Pavel received a gift from Sanaz”. $\dashv$

The proofs of the remaining propositions in this section can be found in Appendix A.

PROPOSITION 4.2.
$z \Vdash \mathsf{H}_s$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{w\}$.

The next proposition shows that Sanaz is happy that Pavel is happy only if she gets the thank-you card and, thus, she knows that he received the gift.

PROPOSITION 4.3.
$z \Vdash \mathsf{H}_s\mathsf{H}_p$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{w\}$.

Note that because Sanaz never acknowledges the thank-you card, Pavel does not know that Sanaz is happy. Hence, he cannot be happy that she is happy. This is captured in the next proposition.

PROPOSITION 4.4.
$z \nVdash \mathsf{H}_p\mathsf{H}_s$(*"Pavel received a gift from Sanaz"*) *for each* $z \in \{w, u, v, t\}$.

PROPOSITION 4.5. $z \nVdash \mathsf{H}_p\mathsf{H}_s\mathsf{H}_p$(*"Pavel received a gift from Sanaz"*) *for each epistemic world* $z \in \{w, u, v, t\}$. $\dashv$

The next proposition states that Sanaz is sad about Pavel not receiving the gift only if she does not send it. Informally, this proposition is true because Sanaz cannot distinguish a world $v$ in which the gift is lost from a world $u$ in which the card is lost.

PROPOSITION 4.6.
$z \Vdash \mathsf{S}_s\neg$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{t\}$.

By Proposition 4.6, Sanaz is sad about Pavel not receiving the gift only if she does not send it. Since Pavel cannot distinguish a world $t$ in which the gift is sent from a world $v$ in which it is lost, Pavel cannot know that Sanaz is sad. This is formally captured in the next proposition.

PROPOSITION 4.7. $z \nVdash \mathsf{K}_p\mathsf{S}_s\neg$(*"Pavel received a gift from Sanaz"*) *for each epistemic world* $z \in \{w, u, t, v\}$.

PROPOSITION 4.8.
$z \Vdash \mathsf{S}_p\neg$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{v, t\}$.

PROPOSITION 4.9.
$z \Vdash \mathsf{S}_s\mathsf{S}_p\neg$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{t\}$.

PROPOSITION 4.10. $z \nVdash \mathsf{K}_p\mathsf{S}_s\mathsf{S}_p\neg$(*"Pavel received a gift from Sanaz"*) *for each epistemic world* $z \in \{w, u, v, t\}$.

| | Iranian | Russian |
|---|---|---|
| Iranian | 1,3 | 0,0 |
| Russian | 0,0 | 3,1 |

Table 1. The Battle of Cuisines. Sanaz is the first player, Pavel is the second.

## 5. The Battle of Cuisines Scenario

In this section, we illustrate Definition 3.1 using a scenario based on a classical strategic game. Suppose that the two co-authors independently decide on a restaurant where to have dinner. Sanaz, being Iranian, wants to explore Russian cuisine, while Pavel would prefer to have dinner in an Iranian restaurant. The epistemic worlds in this model are pairs $(r_s, r_p)$ of restaurant choices made by Sanaz and Pavel respectively, where $r_s, r_p \in \{\text{Iranian}, \text{Russian}\}$. We will consider the situation after they both have already arrived at a restaurant and thus each of them already knows the choice made by the other. Hence, both of them can distinguish all epistemic worlds. In other words, this is a *perfect information* scenario. We specify the preference relations of Sanaz and Pavel through their respective utility functions $u_s$ and $u_p$ captured by the pay-off matrix in Table 1. For example, $(\text{Russian}, \text{Iranian}) \prec_s (\text{Iranian}, \text{Iranian})$ because the value of Sanaz's utility function in the world $(\text{Iranian}, \text{Iranian})$ is larger than in the world $(\text{Russian}, \text{Iranian})$: $u_s(\text{Iranian}, \text{Iranian}) = 1$, $u_s(\text{Russian}, \text{Iranian}) = 0$. The same scenario could also be captured in a diagram depicted in Figure 3.

Proposition 5.1.
$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_p($*"Pavel is in the Russian restaurant"*$)$.

Proof. Note that

$$(\text{Iranian}, \text{Iranian}) \nVdash \text{“Pavel is in the Russian restaurant”},$$
$$(\text{Russian}, \text{Russian}) \Vdash \text{“Pavel is in the Russian restaurant”}.$$

At the same time, see Table 1,

$$u_p(\text{Russian}, \text{Russian}) = 1 < 3 = u_p(\text{Iranian}, \text{Iranian}).$$

Hence, $(\text{Iranian}, \text{Iranian}) \not\prec_p (\text{Russian}, \text{Russian})$. Therefore, by item 6(b) of Definition 3.1,

$$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_p(\text{“Pavel is in the Russian restaurant”}). \quad \dashv$$

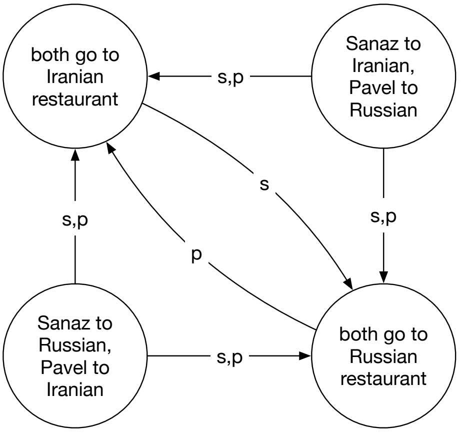


Figure 3. The Battle of Cuisines.

The proofs of the remaining propositions in this section can be found in Appendix B.

Note that Sanaz prefers the world (Russian, Russian) to any other world. This, however, does not mean that she is happy about everything in this world. The next two propositions illustrate this.

PROPOSITION 5.2.
$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_s($*"Sanaz is in the Russian restaurant"*$)$.

PROPOSITION 5.3.
$(x, y) \Vdash \mathsf{H}_s($*"Sanaz and Pavel are in the same restaurant"*$)$ *iff* $x = y$.

PROPOSITION 5.4.
$(x, y) \Vdash \mathsf{H}_p($*"Sanaz and Pavel are in the same restaurant"*$)$ *iff* $x = y$.

PROPOSITION 5.5.
$(x, y) \Vdash \mathsf{H}_p\mathsf{H}_s($*"Sanaz and Pavel are in the same restaurant"*$)$ *iff* $x = y$.

PROPOSITION 5.6.
$(x, y) \Vdash \mathsf{H}_s\mathsf{H}_p\mathsf{H}_s($*"Sanaz and Pavel are in the same restaurant"*$)$ *if and only if* $x = y$.

PROPOSITION 5.7.
$(x, y) \Vdash \mathsf{H}_s($*"Sanaz and Pavel are in the Russian restaurant"*$)$
*if and only if* $x = y = \text{Russian}$.

Proposition 5.8.
$$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_p(\textit{“Sanaz and Pavel are in the Russian restaurant”}).$$

Proposition 5.9.
$$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_p\mathsf{H}_s(\textit{“Sanaz and Pavel are in the Russian restaurant”}).$$

Proposition 5.10.
*$(x, y) \Vdash \mathsf{S}_s$(“Sanaz and Pavel are in different restaurants”) iff $x \neq y$.*

Proposition 5.11.
*$(x, y) \Vdash \mathsf{S}_p$(“Sanaz and Pavel are in different restaurants”) iff $x \neq y$.*

Proposition 5.12.
*$(x, y) \Vdash \mathsf{S}_p\mathsf{S}_s$(“Sanaz and Pavel are in different restaurants”) iff $x \neq y$.*

## 6. The Lottery Scenario

> Be happy for this moment.
> This moment is your life.
>
> The Rubáiyát of Omar Khayyám

As our next example, consider a hypothetical situation where Sanaz and Pavel play lottery with Omar Khayyám, an Iranian mathematician, astronomer, philosopher, and poet. Each of them gets a lottery ticket and it is known that exactly one out of three tickets is the winning ticket. We consider the moment when each of the players has already seen her or his own ticket but does not know yet what are the tickets of the other players. We assume that each of the three players prefers the outcome when she or he wins the lottery.

Figure 4 depicts the epistemic model with preferences capturing the above scenario. It has three epistemic worlds, $u$, $v$, and $w$ in which the winner is Sanaz, Omar, and Pavel respectively. Dashed lines represent indistinguishability relations. For example, the dashed line between worlds $w$ and $v$ labelled with $s$ shows that Sanaz cannot distinguish the world in which Pavel wins from the one in which Omar wins. This is true because we consider the knowledge at the moment when neither of the

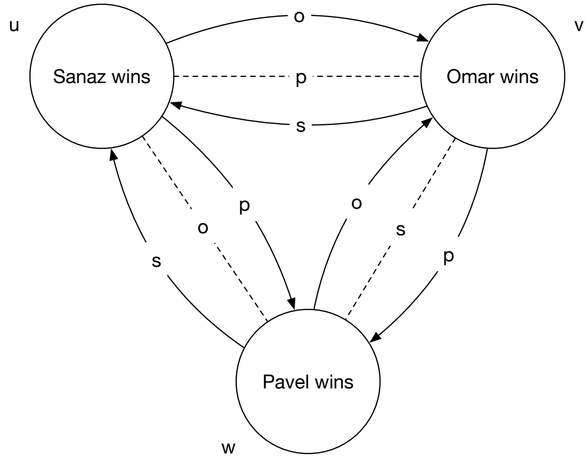


Figure 4. Lottery Epistemic Model with Preferences.

players knows yet what the tickets of the other players are. The directed edges between worlds represent preference relations. For example, the directed edge from $w$ to $v$ labelled with $o$ captures the fact that Omar would prefer to win the lottery rather than lose it.

PROPOSITION 6.1. *$x \Vdash \mathsf{H}_s$("Sanaz won the lottery") iff $x = u$.*

PROOF. ($\Rightarrow$) Suppose that $x \neq u$. Thus, $x \nVdash$ "Sanaz won the lottery", see Figure 4. Therefore, $x \nVdash \mathsf{H}_s$("Sanaz won the lottery") by item 6(a) of Definition 3.1.

($\Leftarrow$) Suppose that $x = u$. To prove the required statement, it suffices to verify conditions (a), (b), and (c) from item 6 of Definition 3.1:

*Condition a*: Consider any world $y$ such that $u \sim_s y$. We will show that $y \Vdash$ "Sanaz won the lottery". Indeed, assumption $u \sim_s y$ implies that $u = y$, see Figure 4. Therefore, see again Figure 4, $y \Vdash$ "Sanaz won the lottery".

*Condition b*: Consider any $y, z$ such that $y \nVdash$ "Sanaz won the lottery" and $z \Vdash$ "Sanaz won the lottery". We will show that $y \prec_s z$. Indeed, by looking at Figure 4, the first assumption implies that $y \in \{w, v\}$. Similarly, the second assumption implies that $z = u$. Statements $y \in \{w, v\}$ and $z = u$ imply that $y \prec_s z$ - see again Figure 4.

*Condition c*: $w \nVdash$ "Sanaz won the lottery". $\dashv$

The proofs of the remaining propositions in this section can be found in Appendix B.1.

Proposition 6.2. $u \nVdash \mathsf{H}_s$(*"Pavel lost the lottery"*).

Proposition 6.3. $u \nVdash \mathsf{K}_p\mathsf{H}_s$(*"Sanaz won the lottery"*).

Proposition 6.4. $u \Vdash \mathsf{S}_p$(*"Pavel lost the lottery"*).

Proposition 6.5. $u \Vdash \mathsf{K}_s\mathsf{S}_p$(*"Pavel lost the lottery"*).

## 7. Undefinability of Sadness through Happiness

In this section, we prove that sadness is not definable through happiness. More formally, we show that the formula $\mathsf{S}_a p$ is not equivalent to any formula in the language $\Phi^{-\mathsf{S}}$:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{N}\varphi \mid \mathsf{K}_a\varphi \mid \mathsf{H}_a\varphi,$$

which is obtained by removing modality $\mathsf{S}$ from the full language $\Phi$ of our logical system. In the next section, we will use a duality principle to claim that happiness is not definable through sadness either.

Without loss of generality, in this section, we assume that the set of agents $\mathcal{A}$ contains a single agent $a$ and the set of propositional variables contains a single variable $p$. To prove the undefinability of sadness through happiness we consider two epistemic models with preferences depicted in Figure 5. By $\Vdash_l$ and $\Vdash_r$ we mean the satisfaction relations for the left and the right model respectively.

In Lemma 7.3, we show that these two models are indistinguishable in language the $\Phi^{-\mathsf{S}}$. In Lemma 7.4 and Lemma 7.5, we prove that $w_1 \Vdash_l \mathsf{S}_a p$ and $w_1 \nVdash_r \mathsf{S}_a p$ respectively. Together, these three statements imply the undefinability of modality $\mathsf{S}$ in the language $\Phi^{-\mathsf{S}}$, which is stated in the end of this section as Theorem 7.1. We start with two auxiliary lemmas used in the proof of Lemma 7.3.

Lemma 7.1. *$w \nVdash_l \mathsf{H}_a\varphi$ for all $w \in \{w_1, w_2, w_3\}$ and $\varphi \in \Phi$.*

Proof. Suppose that $w \Vdash_l \mathsf{H}_a\varphi$. Thus, by item 6 of Definition 3.1,

$$u \Vdash_l \varphi \text{ for each world } u \in W \text{ such that } w \sim_a u, \tag{7.1}$$

$$u \prec_a u' \text{ for all worlds } u, u' \in W \text{ such that } u \nVdash \varphi \text{ and } u' \Vdash \varphi, \tag{7.2}$$

and there is a world $\widehat{w} \in \{w_1, w_2, w_3\}$ such that

$$\widehat{w} \nVdash \varphi. \tag{7.3}$$


Figure 5. Two Models.

Since relation $\sim_a$ is reflexive, statement (7.1) implies that

$$w \Vdash \varphi. \tag{7.4}$$

Thus, using statements (7.2) and (7.3),

$$\widehat{w} \prec_a w. \tag{7.5}$$

Hence, see Figure 5 (left),

$$\widehat{w} \in \{w_1, w_3\} \text{ and } w = w_2. \tag{7.6}$$

Note that $w_2 \sim_a w_3$, see Figure 5 (left). Thus, by (7.6), $w \sim_a w_3$; and, by (7.1):

$$w_3 \Vdash \varphi. \tag{7.7}$$

Hence, $\widehat{w} \neq w_3$ because of statement (7.3). Thus, $\widehat{w} = w_1$ due to (7.6). Then, $w_1 \nVdash \varphi$ because of (7.3). Therefore, $w_1 \prec_a w_3$ by (7.2) and (7.7), which is a contradiction; see Figure 5 (left). ⊣

LEMMA 7.2. *$w \nVdash_r \mathsf{H}_a\varphi$ for all $w \in \{w_1, w_2, w_3\}$ and $\varphi \in \Phi$.*

PROOF. Suppose $w \Vdash_r \mathsf{H}_a\varphi$. Thus, $w \Vdash_r \varphi$ by item 6(a) of Definition 3.1 and the reflexivity of relation $\sim_a$. At the same time, by item 6(c) of Definition 3.1, there must exist a world $u \in W$ such that $u \nVdash_r \varphi$. By item 6(b) of the same Definition 3.1, the assumption $w \Vdash_r \mathsf{H}_a\varphi$ and the statements $u \nVdash_r \varphi$ and $w \Vdash_r \varphi$ imply that $u \prec_a w$, which is a contradiction because relation $\prec_a$ in the right model is empty, see Figure 5. ⊣

LEMMA 7.3. *$w \Vdash_l \varphi$ iff $w \Vdash_r \varphi$ for all $w$ and $\varphi \in \Phi^{-\mathsf{S}}$.*

PROOF. We prove the statement by structural induction on the formula $\varphi$. For the case when $\varphi$ is propositional variable $p$, observe that $\pi_l(p) =$

$\{w_1, w_3\} = \pi_r(p)$ by the choice of the models, see Figure 5. Thus, $w \Vdash_l p$ iff $w \Vdash_r p$ for any world $w$ by item 1 of Definition 3.1. The case when $\varphi$ is a negation or an implication follows from the induction hypothesis and items 2 and 3 of Definition 3.1 in a straightforward way.

Suppose that $\varphi$ has the form $\mathsf{N}\psi$. If $w \nVdash_l \mathsf{N}\psi$, then by item 4 of Definition 3.1, there must exist a world $u \in \{w_1, w_2, w_3\}$ such that $u \nVdash_l \psi$. Hence, by the induction hypothesis, $u \nVdash_r \psi$. Therefore, $w \nVdash_r \mathsf{N}\psi$, by item 4 of Definition 3.1. The proof in the other direction is similar.

Assume that $\varphi$ has the form $\mathsf{K}_a\psi$. If $w \nVdash_l \mathsf{K}_a\psi$, then, by item 5 of Definition 3.1, there is a world $u \in \{w_1, w_2, w_3\}$ such that $w \sim_a^l u$ and $u \nVdash_l \psi$. Then, $w \sim_a^r u$ because relations $\sim^l$ and $\sim^r$ are equal, see Figure 5 and, by the induction hypothesis, $u \nVdash_r \psi$. Therefore, $w \nVdash_r \mathsf{K}_a\psi$ by item 5 of Definition 3.1. The proof in the other direction is similar.

Finally, suppose that $\varphi$ has the form $\mathsf{H}_a\psi$. Therefore, $w \nVdash_l \varphi$ and $w \nVdash_r \varphi$ by Lemma 7.1 and Lemma 7.2 respectively. ⊣

Lemma 7.4. *$w_1 \Vdash_l \mathsf{S}_a p$.*

Proof. We verify the three conditions from item 7 of Definition 3.1 separately:

*Condition a*: Observe that $w_1 \in \{w_1, w_3\} = \pi_l(p)$, see Figure 5 (left). Then, $w_1 \Vdash_l p$ by item 1 of Definition 3.1. Note also that there is only one world $u \in \{w_1, w_2, w_3\}$ such that $w_1 \sim_a^l u$ (namely, the world $w_1$ itself), see Figure 5 (left). Therefore, $u \Vdash p$ for each world $u \in \{w_1, w_2, w_3\}$ such that $w_1 \sim_a u$.

*Condition b*: Note that $\pi_l(p) = \{w_1, w_3\}$, see Figure 5 (left). Then, $w_1 \Vdash_l p$, $w_3 \Vdash_l p$, and $w_2 \nVdash_l p$ by item 1 of Definition 3.1. Also, observe that $w_1 \prec_a^l w_2$ and $w_3 \prec_a^l w_2$, see Figure 5 (left). Thus, for any worlds $u, u' \in \{w_1, w_2, w_3\}$, if $u \Vdash p$ and $u' \nVdash p$, then $u \prec_a u'$.

*Condition c*: $w_2 \nVdash_l p$ by item 1 of Definition 3.1 and because $w_2 \notin \{w_1, w_3\} = \pi_l(p)$, see Figure 5 (left). ⊣

Lemma 7.5. *$w_1 \nVdash_r \mathsf{S}_a p$.*

Proof. Note that $\pi_r(p) = \{w_1, w_3\}$, see Figure 5 (right). Thus, $w_1 \Vdash_r p$ and $w_2 \nVdash_r p$ by item 1 of Definition 3.1. Observe also, that $w_1 \not\prec_a w_2$, Figure 5 (right). Therefore, $w_1 \nVdash_r \mathsf{S}_a p$ by item 7(b) of Definition 3.1. ⊣

The next theorem follows from the three lemmas above.

Theorem 7.1. *Modality $\mathsf{S}$ is not definable in language $\Phi^{-\mathsf{S}}$.* ⊣

## 8. Duality of Happiness and Sadness

As we have shown in the previous section, the sadness modality is not definable through the happiness modality. In spite of this, there still is a connection between these two modalities captured below in Theorem 8.1. To understand this connection, we need to introduce the notion of a converse model and the notion of $\tau$-translation. As usual, for any binary relation $R \subseteq X \times Y$, by converse relation $R^{\mathsf{c}}$ we mean the set of pairs $\{(y,x) \in Y \times X \mid (x,y) \in R\}$.

DEFINITION 8.1. By the converse model $M^{\mathsf{c}}$ of an epistemic model with preference $M = (W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$, we mean the model $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a^{\mathsf{c}}\}_{a\in\mathcal{A}}, \pi)$.

For any epistemic model with preference, by $\Vdash$ we denote the satisfaction relation for this model and by $\Vdash^{\mathsf{c}}$ the satisfaction relation for the converse model.

DEFINITION 8.2. For any formula $\varphi \in \Phi$, the formula $\tau(\varphi) \in \Phi$ is defined recursively as follows:

$\tau(p) = p$, where $p$ is a propositional variable,
$\tau(\neg\varphi) = \neg\tau(\varphi)$,
$\tau(\varphi \to \psi) = \tau(\varphi) \to \tau(\psi)$,
$\tau(\mathsf{N}\varphi) = \mathsf{N}\tau(\varphi)$,
$\tau(\mathsf{K}_a\varphi) = \mathsf{K}_a\tau(\varphi)$,
$\tau(\mathsf{H}_a\varphi) = \mathsf{S}_a\tau(\varphi)$,
$\tau(\mathsf{S}_a\varphi) = \mathsf{H}_a\tau(\varphi)$.

We are now ready to state and prove the "duality principle" that connects modalities $\mathsf{H}$ and $\mathsf{S}$.

THEOREM 8.1. *$w \Vdash \varphi$ iff $w \Vdash^{\mathsf{c}} \tau(\varphi)$, for each world $w$ of an epistemic model with preferences.*

PROOF. We prove the theorem by induction on structural complexity of a formula $\varphi$. If $\varphi$ is a propositional variable, then the statement of the theorem holds because the model and the converse model have the same valuation function $\pi$. If $\varphi$ is a negation, an implication, an $\mathsf{N}$-formula, or an $\mathsf{K}$-formula, then the statement of the theorem follows from the induction hypothesis and items 2–5 of Definition 3.1, respectively.

Suppose that $\varphi$ has the form $\mathsf{H}_a\psi$. First, assume that $w \Vdash \mathsf{H}_a\psi$. Thus, by item 6 of Definition 3.1, the following three conditions are satisfied:

(a) $u \Vdash \psi$ for each world $u \in W$ such that $w \sim_a u$,
(b) for any two worlds $u, u' \in W$, if $u \nVdash \psi$ and $u' \Vdash \psi$, then $u \prec_a u'$,
(c) there is a world $u \in W$ such that $u \nVdash \psi$.

Hence, by the induction hypothesis,

(d) $u \Vdash^{\mathsf{c}} \tau(\psi)$ for each world $u \in W$ such that $w \sim_a u$,
(e) for any two worlds $u, u' \in W$, if $u \nVdash^{\mathsf{c}} \tau(\psi)$ and $u' \Vdash^{\mathsf{c}} \tau(\psi)$, then $u \prec_a u'$,
(f) there is a world $u \in W$ such that $u \nVdash^{\mathsf{c}} \tau(\psi)$.

Note that statement (e) is logically equivalent to

(g) for any two worlds $u, u' \in W$, if $u \Vdash^{\mathsf{c}} \tau(\psi)$ and $u' \nVdash^{\mathsf{c}} \tau(\psi)$, then $u' \prec_a u$.

By the definition of converse partial order, (g) is equivalent to

(h) for any two worlds $u, u' \in W$, if $u \Vdash^{\mathsf{c}} \tau(\psi)$ and $u' \nVdash^{\mathsf{c}} \tau(\psi)$, then $u \prec^{\mathsf{c}}_a u'$.

Thus, $w \Vdash^{\mathsf{c}} \mathsf{S}_a\tau(\psi)$ by item 7 of Definition 3.1 using statements (d), (h), and (f). Therefore, $w \Vdash^{\mathsf{c}} \tau(\mathsf{H}_a\psi)$. The proof in the other direction and the proof for the case when $\varphi$ has the form $\mathsf{S}_a\psi$ are similar. $\dashv$

The next theorem follows from Theorem 7.1 and Theorem 8.1.

Theorem 8.2. *Modality* $\mathsf{H}$ *is not definable in language* $\Phi^{-\mathsf{H}}$.

## 9. Axioms of Emotions

In addition to propositional tautologies in language $\Phi$, our logical system contains the following axioms, where $\mathsf{E} \in \{\mathsf{H}, \mathsf{S}\}$:

Truth: $\mathsf{N}\varphi \to \varphi$, $\mathsf{K}_a\varphi \to \varphi$, and $\mathsf{E}_a\varphi \to \varphi$,
Distributivity:
$\mathsf{N}(\varphi \to \psi) \to (\mathsf{N}\varphi \to \mathsf{N}\psi)$,
$\mathsf{K}_a(\varphi \to \psi) \to (\mathsf{K}_a\varphi \to \mathsf{K}_a\psi)$,
Negative Introspection: $\neg\mathsf{N}\varphi \to \mathsf{N}\neg\mathsf{N}\varphi$, and $\neg\mathsf{K}_a\varphi \to \mathsf{K}_a\neg\mathsf{K}_a\varphi$,
Knowledge of Necessity: $\mathsf{N}\varphi \to \mathsf{K}_a\varphi$,
Emotional Introspection: $\mathsf{E}_a\varphi \to \mathsf{K}_a\mathsf{E}_a\varphi$,
Emotional Consistency: $\mathsf{H}_a\varphi \to \neg\mathsf{S}_a\varphi$,

Coherence of Possible Emotions:
$\overline{\mathsf{N}}\mathsf{E}_a\varphi \wedge \overline{\mathsf{N}}\mathsf{E}_a\psi \to \mathsf{N}(\varphi \to \psi) \vee \mathsf{N}(\psi \to \varphi)$,
$\overline{\mathsf{N}}\mathsf{H}_a\varphi \wedge \overline{\mathsf{N}}\mathsf{S}_a\psi \to \mathsf{N}(\varphi \to \neg\psi) \vee \mathsf{N}(\neg\psi \to \varphi)$,
Counterfactual: $\mathsf{E}_a\varphi \to \neg\mathsf{N}\varphi$,
Emotional Predictability:
$\overline{\mathsf{N}}\mathsf{H}_a\varphi \vee \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \to (\mathsf{K}_a\varphi \to \mathsf{H}_a\varphi)$,
$\overline{\mathsf{N}}\mathsf{H}_a\neg\varphi \vee \overline{\mathsf{N}}\mathsf{S}_a\varphi \to (\mathsf{K}_a\varphi \to \mathsf{S}_a\varphi)$,
Substitution: $\mathsf{N}(\varphi \leftrightarrow \psi) \to (\mathsf{E}_a\varphi \to \mathsf{E}_a\psi)$.

The Truth, the Distributivity, and the Negative Introspection axioms for modalities $\mathsf{N}$ and $\mathsf{K}$ are well-known properties from the modal logic S5. The Truth axiom for modality $\mathsf{E}$ states that if an agent is either happy or sad about $\varphi$, then $\varphi$ must be true. This axiom reflects the fact that our emotions are defined through the agent's knowledge. We will mention belief-based emotions in the conclusion.

The Knowledge of Necessity axioms states that each agent knows all statements that are true in all worlds of the model. The Emotional Introspection axiom captures one of the two possible interpretations of the title of this article: each agent knows her emotions. The other interpretation of the title is stated below as Lemma 9.2. The Emotional Consistency axiom states that an agent cannot be simultaneously happy and sad about the same thing.

Let us now turn our attention to the Coherence of Possible Emotions axioms. Note first that the same agent cannot be happy about statements $\varphi$ and $\neg\varphi$ in the same world because, by item 6(a) of Definition 3.1, that would mean that both of these statements are true in this world. Thus, the formula $\mathsf{H}_a\varphi \wedge \mathsf{H}_a\neg\varphi \to \bot$ is universally true under our semantics. Next, recall that $\overline{\mathsf{N}}\varphi$ stands for the formula $\neg\mathsf{N}\neg\varphi$. Thus, the formula $\overline{\mathsf{N}}\mathsf{H}_a\varphi$ states that an agent $a$ could *possibly* be happy about $\varphi$. Let us now observe that an agent cannot be possibly happy about $\varphi$ and possibly happy about $\neg\varphi$. Indeed, suppose that there are worlds $w_1$ and $w_2$ such that $w_1 \Vdash \mathsf{H}_a\varphi$ and $w_2 \Vdash \mathsf{H}_a\neg\varphi$. Then, $w_1 \Vdash \varphi$ and $w_2 \nVdash \varphi$ by item 6(a) of Definition 3.1. Thus, item 6(b) of Definition 3.1 implies that $w \prec_a u$ and $u \prec_a w$, which contradicts item 3 of Definition 3.1. Hence, the formula $\overline{\mathsf{N}}\mathsf{H}_a\varphi \wedge \overline{\mathsf{N}}\mathsf{H}_a\neg\varphi \to \bot$ is universally true under our semantics. Finally, note that above observation can be generalized to the statement $\overline{\mathsf{N}}\mathsf{H}_a\varphi \wedge \overline{\mathsf{N}}\mathsf{H}_a\psi \to \mathsf{N}(\varphi \to \psi) \vee \mathsf{N}(\psi \to \varphi)$ for arbitrary formulae $\varphi$ and $\psi$. Indeed, suppose that the conclusion of this implication does not hold. Thus, statements $\mathsf{N}(\varphi \to \psi)$ and $\mathsf{N}(\psi \to \varphi)$ are both false. Hence, by

| | Good French Grade | Bad French Grade |
|---|---|---|
| Good Math Grade | 2 | 1 |
| Bad Math Grade | 1 | 0 |

Table 2. Student's Satisfaction Level.

item 4 of Definition 3.1, there must exist worlds $w_1$ and $w_2$ such that $w_1 \nVdash \varphi \to \psi$ and $w_2 \nVdash \psi \to \varphi$. Then, $w_1 \Vdash \varphi \wedge \neg\psi$ and $w_2 \Vdash \psi \wedge \neg\varphi$. Note that $w_1 \Vdash \varphi$ and $w_2 \Vdash \neg\varphi$ imply $w_2 \prec_a w_1$ by the assumption $\overline{\mathsf{N}}\mathsf{H}_a\varphi$ and item 6(b) of Definition 3.1. Similarly, $w_2 \Vdash \psi$ and $w_1 \Vdash \neg\psi$ imply $w_1 \prec_a w_2$ by the assumption $\overline{\mathsf{N}}\mathsf{H}_a\psi$, which is a contradiction. Therefore, the statement $\overline{\mathsf{N}}\mathsf{H}_a\varphi \wedge \overline{\mathsf{N}}\mathsf{H}_a\psi \to \mathsf{N}(\varphi \to \psi) \vee \mathsf{N}(\psi \to \varphi)$ is universally true under our semantics. This statement is a special case of a Coherence of Possible Emotions axiom. In Section 10, we formally prove the soundness of these axioms in the general form along with the soundness of the other axioms of our system. Note that Coherence of Possible Emotions axioms reflect the semantics of happiness and sadness that we proposed in Definition 3.1. As we discuss in the end of Section 12, these axioms do not hold under goodness-based semantics of emotions.

One can express a concern about the Coherence of Possible Emotions axioms, stating that a student is happy to get a good grade on a math exam and is also happy to get a good grade on a French exam, but a good grade on either of these exams seems to imply a good grade on the other. Although at first this appears to be a counterexample to the Coherence of Possible Emotions axioms, it is not. Indeed, consider Table 2 that captures the student's satisfaction level in the described situation. The cells of this table represent epistemic worlds distinguishable to the student. Note that the satisfaction level in the first row of the table is *not* strictly higher than in the second row. Thus, the student is *not* happy to get a good grade on the math exam. Similarly, the student is *not* happy to get a good grade on the French exam. At the same time, the satisfaction level in the world where the student gets at least one good grade is higher than in the world where the student gets no such grades. Thus, the student is happy to have a good grade on *at least one* of these two exams. One can also observe that the student is happy to get good grades *on both* of these exams. Note that getting good grades on both exams implies getting a good grade on at least one exam, just like claimed by the first Coherence of Possible Emotions axiom.

The Counterfactual axiom states that an agent cannot have an emotion about something which is universally true in the model. This axiom reflects items 6(c) and 7(c) of Definition 3.1.

Because the assumptions of both Emotional Predictability axioms contain disjunctions, each of these axioms could be split into two statements. The first statement of the first Emotional Predictability axiom says that if an agent is possibly happy about $\varphi$, then she must be happy about $\varphi$ each time she knows that $\varphi$ is true. The second statement of the same axiom says that if an agent is possibly sad about $\neg\varphi$, then she must be happy about $\varphi$ each time she knows that $\varphi$ is true. The second Emotional Predictability axiom is the dual form of the first axiom.

Finally, the Substitution axiom states that if two statements are equivalent in each world of the model and an agent has an emotion about one of them, then she must have the same emotion about the other statement.

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

$$\frac{\varphi, \varphi \to \psi}{\psi} \qquad\qquad \frac{\varphi}{\mathsf{N}\varphi}.$$

For any set of statements $X \subseteq \Phi$, we write $X \vdash \varphi$ if a formula $\varphi$ is derivable from the theorems of our system and a set $X$ using only the Modus Ponens inference rule. We say that a set $X$ is *inconsistent* if there is a formula $\varphi \in \Phi$ such that $X \vdash \varphi$ and $X \vdash \neg\varphi$.

LEMMA 9.1. *Inference rule* $\dfrac{\varphi}{\mathsf{K}_a\varphi}$ *is derivable.*

PROOF. This rule is a combination of the Necessitation rule, the Knowledge of Necessity axiom, and Modus Ponens. $\dashv$

LEMMA 9.2. $\vdash \mathsf{E}_a\varphi \to \mathsf{K}_a\varphi$.

PROOF. Note that $\vdash \mathsf{E}_a\varphi \to \varphi$ by the Truth axiom. Thus, $\vdash \mathsf{K}_a(\mathsf{E}_a\varphi \to \varphi)$ by Lemma 9.1. Hence, $\vdash \mathsf{K}_a\mathsf{E}_a\varphi \to \mathsf{K}_a\varphi$ by the Distributivity axiom and the Modus Ponens inference rule. Therefore, $\vdash \mathsf{E}_a\varphi \to \mathsf{K}_a\varphi$ by the Emotional Introspection axiom and propositional reasoning. $\dashv$

The next three lemmas are well known in model logic. We omit their proofs.

LEMMA 9.3 (deduction). *If* $X, \varphi \vdash \psi$, *then* $X \vdash \varphi \to \psi$.

LEMMA 9.4. *If* $\varphi_1, \dots, \varphi_n \vdash \psi$, *then* $\Box\varphi_1, \dots, \Box\varphi_n \vdash \Box\psi$, *where* $\Box$ *is either modality* $\mathsf{N}$ *or modality* $\mathsf{K}_a$.

LEMMA 9.5 (positive introspection). $\vdash \Box\varphi \to \Box\Box\varphi$, *where* $\Box$ *is either modality* $\mathsf{N}$ *or modality* $\mathsf{K}_a$.

LEMMA 9.6. 1. $\mathsf{N}(\varphi \to \psi), \mathsf{N}(\neg\varphi \to \neg\psi) \vdash \mathsf{N}(\varphi \leftrightarrow \psi)$,
2. $\mathsf{N}(\varphi \to \neg\psi), \mathsf{N}(\neg\varphi \to \psi) \vdash \mathsf{N}(\varphi \leftrightarrow \neg\psi)$.

PROOF. It is provable in the propositional logic that $\varphi \to \psi, \neg\varphi \to \neg\psi \vdash \varphi \leftrightarrow \psi$. Thus, $\mathsf{N}(\varphi \to \psi), \mathsf{N}(\neg\varphi \to \neg\psi) \vdash \mathsf{N}(\varphi \leftrightarrow \psi)$ by Lemma 9.4. The proof of the second part of the lemma is similar. ⊣

LEMMA 9.7. 1. $\mathsf{N}(\varphi \leftrightarrow \psi), \overline{\mathsf{N}}\mathsf{H}_a\varphi \vdash \overline{\mathsf{N}}\mathsf{H}_a\psi$,
2. $\mathsf{N}(\varphi \leftrightarrow \psi), \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \vdash \overline{\mathsf{N}}\mathsf{S}_a\neg\psi$,
3. $\mathsf{N}(\varphi \leftrightarrow \neg\psi), \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \vdash \overline{\mathsf{N}}\mathsf{S}_a\psi$.

PROOF. Formula $\mathsf{N}(\varphi \leftrightarrow \psi) \to (\mathsf{H}_a\varphi \to \mathsf{H}_a\psi)$ is an instance of the Substitution axiom. Thus, $\vdash \mathsf{N}(\varphi \leftrightarrow \psi) \to (\neg\mathsf{H}_a\psi \to \neg\mathsf{H}_a\varphi)$ by the laws of propositional reasoning. Hence, $\mathsf{N}(\varphi \leftrightarrow \psi), \neg\mathsf{H}_a\psi \vdash \neg\mathsf{H}_a\varphi$ by the Modus Ponens rule applied twice. Then, $\mathsf{N}\mathsf{N}(\varphi \leftrightarrow \psi), \mathsf{N}\neg\mathsf{H}_a\psi \vdash \mathsf{N}\neg\mathsf{H}_a\varphi$ by Lemma 9.4. Thus, by Lemma 9.5 and the Modus Ponens inference rule, $\mathsf{N}(\varphi \leftrightarrow \psi), \mathsf{N}\neg\mathsf{H}_a\psi \vdash \mathsf{N}\neg\mathsf{H}_a\varphi$. Hence, $\mathsf{N}(\varphi \leftrightarrow \psi) \vdash \mathsf{N}\neg\mathsf{H}_a\psi \to \mathsf{N}\neg\mathsf{H}_a\varphi$ by Lemma 9.3. Then, by the laws of propositional reasoning, $\mathsf{N}(\varphi \leftrightarrow \psi) \vdash \neg\mathsf{N}\neg\mathsf{H}_a\varphi \to \neg\mathsf{N}\neg\mathsf{H}_a\psi$. Thus, by the definition of modality $\overline{\mathsf{N}}$, we have $\mathsf{N}(\varphi \leftrightarrow \psi) \vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi \to \overline{\mathsf{N}}\mathsf{H}_a\psi$. Therefore, by the Modus Ponens inference rule, $\mathsf{N}(\varphi \leftrightarrow \psi), \overline{\mathsf{N}}\mathsf{H}_a\varphi \vdash \overline{\mathsf{N}}\mathsf{H}_a\psi$.

To prove the second statement, observe that $(\varphi \leftrightarrow \psi) \to (\neg\varphi \leftrightarrow \neg\psi)$ is a propositional tautology. Thus, $\varphi \leftrightarrow \psi \vdash \neg\varphi \leftrightarrow \neg\psi$ by Modus Ponens. Hence, $\mathsf{N}(\varphi \leftrightarrow \psi) \vdash \mathsf{N}(\neg\varphi \leftrightarrow \neg\psi)$ by Lemma 9.4. Then, to prove the second statement, it suffices to show that $\mathsf{N}(\neg\varphi \leftrightarrow \neg\psi), \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \vdash \overline{\mathsf{N}}\mathsf{S}_a\neg\psi$. The proof of this is the same as the proof of the first statement.

The proof of the third statement is similar to the proof of the second, but it starts with the tautology $(\varphi \leftrightarrow \neg\psi) \to (\neg\varphi \leftrightarrow \psi)$. ⊣

LEMMA 9.8 (Lindenbaum). *Any consistent set of formulae can be extended to a maximal consistent set of formulae.*

PROOF. The standard proof of Lindenbaum's lemma applies here [12, Proposition 2.14]. ⊣

## 10. Soundness

The Truth, the Distributivity, and the Negative Introspection axioms for modalities $\mathsf{K}$ and $\mathsf{N}$ are well-known principles of S5 logic. The soundness of the Knowledge of Necessity axiom follows from Definition 3.1. Below we prove the soundness of each of the remaining axioms as a separate lemma. We state strong soundness as Theorem 10.1 at the end of the section. In the lemmas below we assume that $w \in W$ is an arbitrary world of an epistemic model with preferences $M = (W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$.

LEMMA 10.1. *If $w \Vdash \mathsf{E}_a\varphi$, then $w \Vdash \varphi$.*

PROOF. First, we consider the case $\mathsf{E} = \mathsf{H}$. Note that $w \sim_a w$ because $\sim_a$ is an equivalence relation. Thus, the assumption $w \Vdash \mathsf{H}_a\varphi$ implies $w \Vdash \varphi$ by item 6(a) of Definition 3.1. The proof for the case $\mathsf{E} = \mathsf{S}$ is similar, but it uses item 7(a) of Definition 3.1 instead of item 6(a). ⊣

LEMMA 10.2. *If $w \Vdash \mathsf{E}_a\varphi$, then $w \Vdash \mathsf{K}_a\mathsf{E}_a\varphi$.*

PROOF. First, we consider the case $\mathsf{E} = \mathsf{H}$. Consider any world $w' \in W$ such that $w \sim_a w'$. By item 5 of Definition 3.1, it suffices to show that $w' \Vdash \mathsf{H}_a\varphi$. Indeed, by item 6 of Definition 3.1, the assumption $w \Vdash \mathsf{H}_a\varphi$ of the lemma implies that

(a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
(b) for any two worlds $u, u' \in W$, if $u \nVdash \varphi$ and $u' \Vdash \varphi$, then $u \prec_a u'$,
(c) there is a world $u \in W$ such that $u \nVdash \varphi$,

By the assumption $w \sim_a w'$, statement (a) implies that

(a′) $u \Vdash \varphi$ for each world $u \in W$ such that $w' \sim_a u$.

Finally, statements (a′), (b), and (c) imply that $w' \Vdash \mathsf{H}_a\varphi$ by item 6 of Definition 3.1. The proof for the case $\mathsf{E} = \mathsf{S}$ is similar, but it uses item 7 of Definition 3.1 instead of item 6. ⊣

LEMMA 10.3. *If $w \Vdash \mathsf{H}_a\varphi$, then $w \nVdash \mathsf{S}_a\varphi$.*

PROOF. By item 6(a) of Definition 3.1, the assumption $w \Vdash \mathsf{H}_a\varphi$ implies that $w \Vdash \varphi$. By item 6(c) of Definition 3.1, the same assumption implies that there is a world $w' \in W$ such that $w' \nVdash \varphi$. By item 6(b) of Definition 3.1 the assumption $w \Vdash \mathsf{H}_a\varphi$ and statements $w' \nVdash \varphi$ and $w \Vdash \varphi$ imply that $w' \prec_a w$. Thus, $w \not\prec_a w'$ because relation $\prec_a$ is a strict partial order. Therefore, $w \nVdash \mathsf{S}_a\varphi$ by item 7(b) of Definition 3.1 and statements $w \Vdash \varphi$ and $w' \nVdash \varphi$. ⊣

LEMMA 10.4. *For any* $\mathsf{E} \in \{\mathsf{H}, \mathsf{S}\}$, *if* $w \Vdash \overline{\mathsf{N}}\mathsf{E}_a\varphi$ *and* $w \Vdash \overline{\mathsf{N}}\mathsf{E}_a\psi$, *then either* $w \Vdash \mathsf{N}(\varphi \to \psi)$ *or* $w \Vdash \mathsf{N}(\psi \to \varphi)$.

PROOF. First, we consider the case $\mathsf{E} = \mathsf{H}$. Suppose that $w \nVdash \mathsf{N}(\varphi \to \psi)$ and $w \nVdash \mathsf{N}(\psi \to \varphi)$. Thus, by item 4 of Definition 3.1, there are epistemic worlds $w_1, w_2 \in W$, such that $w_1 \nVdash \varphi \to \psi$ and $w_2 \nVdash \psi \to \varphi$. Hence, by item 3 of Definition 3.1,

$$w_1 \Vdash \varphi, \quad w_1 \nVdash \psi, \quad w_2 \Vdash \psi, \quad w_2 \nVdash \varphi. \tag{10.1}$$

At the same time, by the definition of modality $\overline{\mathsf{N}}$ and items 2 and 4 of Definition 3.1, the assumption $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$ of the lemma implies that there is a world $w'$ such that $w' \Vdash \mathsf{H}_a\varphi$. Hence, $w_2 \prec_a w_1$ by item 6(b) of Definition 3.1 and parts $w_2 \nVdash \varphi$ and $w_1 \Vdash \varphi$ of statement (10.1).

Similarly, the assumption $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\psi$ of the lemma and parts $w_1 \nVdash \psi$ and $w_2 \Vdash \psi$ of statement (10.1) imply that $w_1 \prec_a w_2$. Note that statements $w_2 \prec_a w_1$ and $w_1 \prec_a w_2$ are inconsistent because relation $\prec_a$ is a strict partial order.

The proof in the case $\mathsf{E} = \mathsf{S}$ is similar, but it uses item 7(b) of Definition 3.1 instead of item 6(b). ⊣

LEMMA 10.5. *If* $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$ *and* $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\psi$, *then either* $w \Vdash \mathsf{N}(\varphi \to \neg\psi)$ *or* $w \Vdash \mathsf{N}(\neg\psi \to \varphi)$.

PROOF. Suppose that $w \nVdash \mathsf{N}(\varphi \to \neg\psi)$ and $w \nVdash \mathsf{N}(\neg\psi \to \varphi)$. Thus, by item 4 of Definition 3.1, there are epistemic worlds $w_1, w_2 \in W$, such that $w_1 \nVdash \varphi \to \neg\psi$ and $w_2 \nVdash \neg\psi \to \varphi$. Hence, by item 3 and item 2 of Definition 3.1,

$$w_1 \Vdash \varphi, \quad w_1 \Vdash \psi, \quad w_2 \nVdash \psi, \quad w_2 \nVdash \varphi. \tag{10.2}$$

At the same time, by the definition of modality $\overline{\mathsf{N}}$ and items 2 and 4 of Definition 3.1, the assumption $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$ of the lemma implies that there is a world $w'$ such that $w' \Vdash \mathsf{H}_a\varphi$. Hence, $w_2 \prec_a w_1$ by item 6(b) of Definition 3.1 and parts $w_2 \nVdash \varphi$ and $w_1 \Vdash \varphi$ of statement (10.2).

Also, by item 7(b) of Definition 3.1, the assumption $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\psi$ of the lemma and parts $w_1 \Vdash \psi$ and $w_2 \nVdash \psi$ of statement (10.2) imply that $w_1 \prec_a w_2$. Note that statements $w_2 \prec_a w_1$ and $w_1 \prec_a w_2$ are inconsistent because relation $\prec_a$ is a strict partial order. ⊣

LEMMA 10.6. *If* $w \Vdash \mathsf{E}_a\varphi$, *then* $w \nVdash \mathsf{N}\varphi$.

PROOF. First, suppose that $\mathsf{E} = \mathsf{H}$. Then, by item 6(c) of Definition 3.1, the assumption $w \Vdash \mathsf{E}_a\varphi$ implies that there is an epistemic world $u \in W$ such that $u \nVdash \varphi$. Therefore, $w \nVdash \mathsf{N}\varphi$ by item 4 of Definition 3.1.

The proof in the case $\mathsf{E} = \mathsf{S}$ is similar, but it uses item 7(c) of Definition 3.1 instead of item 6(c). $\dashv$

LEMMA 10.7. *If either $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$ or $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi$, then the statement $w \Vdash \mathsf{K}_a\varphi$ implies $w \Vdash \mathsf{H}_a\varphi$.*

PROOF. First, suppose that $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$. Thus, by the definition of modality $\overline{\mathsf{N}}$ and items 2 and 4 of Definition 3.1, there is an epistemic world $w' \in W$ such that $w' \Vdash \mathsf{H}_a\varphi$. Hence, by item 6 of Definition 3.1,

(a) $u \Vdash \varphi$ for each world $u \in W$ such that $w' \sim_a u$,
(b) for any two worlds $u, u' \in W$, if $u \nVdash \varphi$ and $u' \Vdash \varphi$, then $u \prec_a u'$,
(c) there is a world $u \in W$ such that $u \nVdash \varphi$.

Also, by item 5 of Definition 3.1, the assumption $w \Vdash \mathsf{K}_a\varphi$ implies that

(a′) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$.

By item 6 of Definition 3.1, statements (a′), (b), and (c) imply $w \Vdash \mathsf{H}_a\varphi$.

Next, suppose that $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi$. Thus, by the definition of modality $\overline{\mathsf{N}}$ and items 2 and 4 of Definition 3.1, there is an epistemic world $w' \in W$ such that $w' \Vdash \mathsf{S}_a\neg\varphi$. Hence, by item 7 of Definition 3.1,

(a) $u \Vdash \neg\varphi$ for each world $u \in W$ such that $w' \sim_a u$,
(b) for any two worlds $u, u' \in W$, if $u \Vdash \neg\varphi$ and $u' \nVdash \neg\varphi$, then $u \prec_a u'$,
(c) there is a world $u \in W$ such that $u \nVdash \neg\varphi$.

Also, by item 5 of Definition 3.1, the assumption $w \Vdash \mathsf{K}_a\varphi$ implies that

(a′) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$.

Note that by item 2 of Definition 3.1, statement (b) implies that

(b′) for any two worlds $u, u' \in W$, if $u \nVdash \varphi$ and $u' \Vdash \varphi$, then $u \prec_a u'$.

And, by item 2 of Definition 3.1, statement (a) implies that

(c′) $w' \nVdash \varphi$

because relation $\sim_a$ is reflexive. Finally, note that by item 6 of Definition 3.1, statements (a′), (b′), and (c′) imply $w \Vdash \mathsf{H}_a\varphi$. $\dashv$

The proof of the next lemma uses the converse models and translation $\tau$ that have been introduced in Definition 8.1 and Definition 8.2 respectively.

Lemma 10.8. *If either* $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\neg\varphi$ *or* $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\varphi$*, then the statement* $w \Vdash \mathsf{K}_a\varphi$ *implies* $w \Vdash \mathsf{S}_a\varphi$*.*

Proof. Let $M^{\mathsf{c}}$ be the converse model of the epistemic model with preferences $M$ and $\Vdash^{\mathsf{c}}$ be the satisfaction relation for the model $M^{\mathsf{c}}$. By Lemma 10.7, if either $w \Vdash^{\mathsf{c}} \overline{\mathsf{N}}\mathsf{H}_a\tau(\varphi)$ or $w \Vdash^{c} \overline{\mathsf{N}}\mathsf{S}_a\neg\tau(\varphi)$, then the statement $w \Vdash^{c} \mathsf{K}_a\tau(\varphi)$ implies $w \Vdash^{\mathsf{c}} \mathsf{H}_a\tau(\varphi)$. Thus, by Definition 8.2, if either $w \Vdash^{\mathsf{c}} \tau(\overline{\mathsf{N}}\mathsf{S}_a\varphi)$ or $w \Vdash^{c} \tau(\overline{\mathsf{N}}\mathsf{H}_a\neg\varphi)$, then the statement $w \Vdash^{c} \tau(\mathsf{K}_a\varphi)$ implies $w \Vdash^{\mathsf{c}} \tau(\mathsf{S}_a\varphi)$. Therefore, by Theorem 8.1, if either $w \Vdash \overline{\mathsf{N}}\mathsf{S}_a\varphi$ or $w \Vdash \overline{\mathsf{N}}\mathsf{H}_a\neg\varphi$, then the statement $w \Vdash \mathsf{K}_a\varphi$ implies $w \Vdash \mathsf{S}_a\varphi$. ⊣

Lemma 10.9. *If* $w \Vdash \mathsf{N}(\varphi \leftrightarrow \psi)$ *and* $w \Vdash \mathsf{E}_a\varphi$*, then* $w \Vdash \mathsf{E}_a\psi$*.*

Proof. First, we consider the case $\mathsf{E} = \mathsf{H}$. By item 6 of Definition 3.1, the assumption $w \Vdash \mathsf{H}_a\varphi$ implies that

(a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
(b) for any two worlds $u, u' \in W$, if $u \nVdash \varphi$ and $u' \Vdash \varphi$, then $u \prec_a u'$,
(c) there is a world $u \in W$ such that $u \nVdash \varphi$.

Thus, by the assumption $w \Vdash \mathsf{N}(\varphi \leftrightarrow \psi)$ and item 4 of Definition 3.1,

(a′) $u \Vdash \psi$ for each world $u \in W$ such that $w \sim_a u$,
(b′) for any two worlds $u, u' \in W$, if $u \nVdash \psi$ and $u' \Vdash \psi$, then $u \prec_a u'$,
(c′) there is a world $u \in W$ such that $u \nVdash \psi$.

By item 6 of Definition 3.1, statements (a′), (b′), and (c′) imply $w \Vdash \mathsf{H}_a\psi$.

The proof in the case $\mathsf{E} = \mathsf{S}$ is similar, but it uses item 7 of Definition 3.1 instead of item 6. ⊣

The strong soundness theorem below follows from the lemmas proven above.

Theorem 10.1. *For any epistemic 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* $w \Vdash \varphi$*.* ⊣

## 11. Utilitarian Emotions

Lang, van er Torre, and Weydert introduced a notion of utilitarian desire which is based on a utility function rather than a preference relation [9]. Although desire, as an emotion, is different from the happiness and sadness emotions that we study in this article, their approach could be adapted to happiness and sadness as well. To do this, one needs to modify Definition 2.1 to include agent-specific utility functions instead of agent-specific preference relations:

DEFINITION 11.1. A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{u_a\}_{a\in\mathcal{A}}, \pi)$ is called an epistemic model with utilities 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. $u_a$ is a "utility" function from set $W$ to real numbers for each agent $a \in \mathcal{A}$,
4. $\pi(p)$ is a subset of $W$ for each propositional variable $p$.

Below is the definition of the satisfaction relation for the epistemic model with utilities. Its parts 6(b) and 7(b) are similar to the utilitarian desire definition in [9]. Unlike the current article, [9] does not prove any completeness results. In the definition below we assume that language $\Phi$ is modified to incorporate a no-negative real "degree" parameter into modalities $\mathsf{H}^d_a$ and $\mathsf{S}^d_a$. We read the statement $\mathsf{H}^d_a\varphi$ as "an agent $a$ is happy about $\varphi$ with degree $d$". Similarly, we read $\mathsf{S}^d_a\varphi$ as "an agent $a$ is sad about $\varphi$ with degree $d$".

DEFINITION 11.2. For any model with utilities $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{u_a\}_{a\in\mathcal{A}}, \pi)$, any world $w \in W$, and any formula $\varphi \in \Phi$, satisfaction relation $w \Vdash \varphi$ is defined 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. $w \Vdash \mathsf{N}\varphi$, if $v \Vdash \varphi$ for each world $v \in W$,
5. $w \Vdash \mathsf{K}_a\varphi$, if $v \Vdash \varphi$ for each world $v \in W$ such that $w \sim_a v$,
6. $w \Vdash \mathsf{H}^d_a\varphi$, if the following three conditions are satisfied:
   (a) $v \Vdash \varphi$ for each world $v \in W$ such that $w \sim_a v$,
   (b) for any $v, v' \in W$, if $v \nVdash \varphi$ and $v' \Vdash \varphi$, then $u_a(v) + d \leq u_a(v')$,
   (c) there is a world $v \in W$ such that $v \nVdash \varphi$,
7. $w \Vdash \mathsf{S}^d_a\varphi$, if the following three conditions are satisfied:
   (a) $v \Vdash \varphi$ for each world $v \in W$ such that $w \sim_a u$,
   (b) for any $v, v' \in W$, if $v \Vdash \varphi$ and $v' \nVdash \varphi$, then $u_a(v) + d \leq u_a(v')$,
   (c) there is a world $u \in W$ such that $u \nVdash \varphi$.

We have already defined utility functions for our Battle of Cuisines scenario, see Section 5. In the two propositions below we use this scenario to illustrate the utilitarian happiness modality. Note how Sanaz is much happier to be with Pavel in a Russian restaurant than she is to be with him in a restaurant.

Proposition 11.1. $(x, x) \Vdash \mathsf{H}_s^1$(*"Sanaz and Pavel are in the same restaurant"*), *where* $x \in \{Iranian, Russian\}$.

Proof. We verify conditions (a)–(c) from item 6 of Definition 11.2.

*Condition a*: Consider any epistemic world $(y, z)$ such that $(x, x) \sim_s (y, z)$. It suffices to show that

$$(y, z) \Vdash \text{"Sanaz and Pavel are in the same restaurant"}.$$

The latter is true because assumption $(x, x) \sim_s (y, z)$ implies that $x = y$ and $x = z$ in the perfect information setting of the Battle of Cuisines scenario.

*Condition b*: Consider any worlds $(y, z)$ and $(y', z')$ such that

$$(y, z) \nVdash \text{"Sanaz and Pavel are in the same restaurant"}, \tag{11.1}$$

$$(y', z') \Vdash \text{"Sanaz and Pavel are in the same restaurant"}. \tag{11.2}$$

Statement (11.1) implies that $u_s(y, z) = 0$, see Table 1. Similarly, statement (11.2) implies that $u_s(y', z') \geq 1$. Therefore, $u_s(y, z) + 1 = 1 \leq u_s(y', z')$.

*Condition c*:

$$(\text{Russian}, \text{Iranian}) \nVdash \text{"Sanaz and Pavel are in the same restaurant"}. \quad \dashv$$

Proposition 11.2.

$$(\text{Russian}, \text{Russian}) \Vdash \mathsf{H}_s^2(\textit{"Sanaz and Pavel are in the Russian restaurant"}).$$

Proof. Conditions (a) and (c) from item 6 of Definition 11.2 could be verified similarly to the proof of Proposition 11.1. Below we verify condition (b).

Consider any worlds $(y, z)$ and $(y', z')$ such that

$$(y, z) \nVdash \text{"Sanaz and Pavel are in the Russian restaurant"}, \tag{11.3}$$

$$(y', z') \Vdash \text{"Sanaz and Pavel are in the Russian restaurant"}. \tag{11.4}$$

Statement (11.3) implies that $u_s(y, z) \leq 1$, see Table 1. Similarly, statement (11.4) implies that $u_s(y', z') = 3$. Therefore, $u_s(y, z) + 2 \leq 1 + 2 = 3 = u_s(y', z')$. $\dashv$

## 12. Goodness-Based Emotions

Lorini and Schwarzentruber proposed a different framework for defining emotions [11]. Instead of specifying preference relations on the epistemic worlds, they label some of the worlds as desirable or "good". In such a setting they define modalities "rejoice" and "disappointment" that are similar to our modalities "happiness" and "sadness". In this section, we compare their approach to ours. Although their framework endows agents with actions, it appears that actions are essential for defining regret and are less important for capturing rejoicing and disappointment. In the definition below, we simplify Lorini and Schwarzentruber's framework to action-less models that we call epistemic models with goodness.

DEFINITION 12.1. A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{G_a\}_{a\in\mathcal{A}}, \pi)$ is called an epistemic model with goodness 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. $G_a \subseteq W$ is a nonempty set of "good" epistemic worlds for an agent $a \in \mathcal{A}$,
4. $\pi(p)$ is a subset of $W$ for each propositional variable $p$.

To represent the gift example from Figure 1 as an epistemic model with goodness, we need to specify the sets of good epistemic worlds $G_s$ and $G_p$ of Sanaz and Pavel. A natural way to do this is to assume that $G_s = G_p = \{w\}$. In other words, the ideal outcome for both of them would be if the gift and the card reached the recipients.

In the lottery example, the desirable outcome for each agent is when the agent wins the lottery. In other words, $G_s = \{u\}$, $G_p = \{w\}$, and $G_o = \{v\}$, see Figure 4.

In the Battle of Cuisines example captured in Table 1, the choice of good epistemic worlds is not obvious. On one hand, we can assume that good worlds for both Sanaz and Pavel are the ones where they have positive pay-offs. In this case,

$$G_s = G_p = \{(\mathrm{Iranian}, \mathrm{Iranian}), (\mathrm{Russian}, \mathrm{Russian})\}.$$

Alternatively, we can choose the good worlds to be those where they get the maximal payoff. In that case, $G_s = \{(\mathrm{Russian}, \mathrm{Russian})\}$ and $G_p = \{(\mathrm{Iranian}, \mathrm{Iranian})\}$. Note that our epistemic models with preferences approach provides a more fine-grained semantics that does not force the choice between these two alternatives.

In the definition below, we rephrase Lorini and Schwarzentruber's formal definitions of "rejoice" and "disappointment" in terms of epistemic models with goodness. We denote modalities "rejoice" and "disappointment" by $\mathsf{H}$ and $\mathsf{S}$ respectively to be consistent with the notations in the rest of this article.

DEFINITION 12.2. For any world $w \in W$ of any epistemic model with goodness $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{G_a\}_{a\in\mathcal{A}}, \pi)$ and any formula $\varphi \in \Phi$, the satisfaction relation $w \Vdash \varphi$ is defined 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. $w \Vdash \mathsf{N}\varphi$, if $u \Vdash \varphi$ for each world $u \in W$,
5. $w \Vdash \mathsf{K}_a\varphi$, if $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
6. $w \Vdash \mathsf{H}_a\varphi$, if the following three conditions are satisfied:
   (a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
   (b) $u \Vdash \varphi$ for each world $u \in G_a$,
   (c) there is a world $u \in W$ such that $u \nVdash \varphi$,
7. $w \Vdash \mathsf{S}_a\varphi$, if the following three conditions are satisfied:
   (a) $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_a u$,
   (b) $u \nVdash \varphi$ for each world $u \in G_a$,
   (c) there is a world $u \in W$ such that $u \nVdash \varphi$.

Consider the discussed above epistemic model with goodness for the gift scenario in which $G_s = G_p = \{w\}$. It is relatively easy to see that all propositions that we proved in Section 4 for the preference-based semantics hold true under the goodness-based semantics of modalities $\mathsf{H}$ and $\mathsf{S}$ given in Definition 12.2.

The situation is different for the Battle of Cuisines scenario. If $\Vdash_1$ denotes the satisfaction relation of the epistemic model with goodness where

$$G_s = G_p = \{(\text{Iranian}, \text{Iranian}), (\text{Russian}, \text{Russian})\},$$

then the following two propositions are true just like they are under our definition of happiness (see Proposition 5.3 and Proposition 5.2):

PROPOSITION 12.1.

$$(\text{Russian}, \text{Russian}) \Vdash_1 \mathsf{H}_s(\textit{"Sanaz and Pavel are in the same restaurant"}).$$

Proof. It suffices to verify conditions (a), (b) and (c) of item 6 from Definition 12.2.

*Condition a*: Since the Battle of the Cuisines is a setting with perfect information, it suffices to show that

$$(\text{Russian}, \text{Russian}) \Vdash_1 (\text{“Sanaz and Pavel are in the same restaurant”}).$$

Note that the last statement is true by the definition of the world (Russian, Russian).

*Condition b*: Statement “Sanaz and Pavel are in the same restaurant” is satisfied in both good worlds: (Iranian, Iranian) and (Russian, Russian).

*Condition* : Statement “Sanaz and Pavel are in the same restaurant” is not satisfied in the world (Russian, Iranian). ⊣

Proposition 12.2.

$$(\text{Russian}, \text{Russian}) \nVdash_1 \mathsf{H}_s(\textit{“Sanaz is in the Russian restaurant”}).$$

Proof. Note that (Iranian, Iranian) is a good world, in which the statement “Sanaz is in the Russian restaurant” is not satisfied. Therefore, the statement of the proposition is true by item 6(b) of Definition 12.2. ⊣

However, for the relation $\Vdash_2$ of the epistemic model with goodness where $G_s = \{(\text{Russian}, \text{Russian})\}$ and $G_p = \{(\text{Iranian}, \text{Iranian})\}$, the situation is different:

Proposition 12.3.

$$(\text{Russian}, \text{Russian}) \Vdash_2 \mathsf{H}_s(\textit{“Sanaz and Pavel are in the same restaurant”}).$$

Proof. The proof is similar to the proof of Proposition 12.1 except that in *Condition b* we only need to consider the world (Russian, Russian). ⊣

Proposition 12.4.

$$(\text{Russian}, \text{Russian}) \Vdash_2 \mathsf{H}_s(\textit{“Sanaz is in the Russian restaurant”}).$$

Proof. It suffices to verify conditions (a), (b), and (c) of item 6 of Definition 12.2:

*Condition a*: Since the Battle of the Cuisines is a setting with perfect information, it suffices to show that

$$(\text{Russian}, \text{Russian}) \Vdash_2 (\text{“Sanaz is in the Russian restaurant”}).$$

Observe that the last statement is true by the definition of the world (Russian, Russian).

*Condition b*: Note that in the current setting set $G_s$ contains only element (Russian, Russian) and that

$$\text{(Russian, Russian)} \Vdash_2 (\text{“Sanaz is in the Russian restaurant”}).$$

*Condition c*:

$$\text{(Iranian, Russian)} \nVdash_2 (\text{“Sanaz is in the Russian restaurant”}). \qquad \dashv$$

We conclude this section by an observation that the Coherence of Possible Emotions axiom is not universally true under the goodness-based semantics. Indeed, note that according to Propositions 12.3 and 12.4, there is an epistemic world in which Sanaz is happy that “Sanaz is in the Russian restaurant” and there is an epistemic world in which she is happy that “Sanaz and Pavel are in the same restaurant”. If the Coherence of Possible Emotions axiom holds in this setting, then one of these statements would imply the other, but neither of them does.

## 13. Canonical Model

In the rest of this article, we prove the strong completeness of our logical system with respect to the semantics given in Definition 3.1. As usual, the proof of the completeness is based on a construction of a canonical model. In this section, for any maximal consistent set of formulae $X_0 \subseteq \Phi$, we define a canonical epistemic model with preferences $M(X_0) = (W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a\}_{a\in\mathcal{A}}, \pi)$.

As common in modal logic, we define worlds as maximal consistent sets of formulae. Since the meaning of modality $\mathsf{N}$ in our system is “for all worlds”, we require all worlds in the canonical model to have the same $\mathsf{N}$-formulae. We achieve this through the following definition.

Definition 13.1. *$W$ is the set of all such maximal consistent sets of formulae $Y$ that $\{\varphi \in \Phi \mid \mathsf{N}\varphi \in X_0\} \subseteq Y$.*

Note that although the above definition only requires all $\mathsf{N}$-formulae from set $X_0$ to be in set $Y$, it is possible to show that the converse is also true due to the presence of the Negative Introspection axiom for modality $\mathsf{N}$ in our system.

LEMMA 13.1. $X_0 \in W$.

PROOF. Consider any formula $\mathsf{N}\varphi \in X_0$. By Definition 13.1, it suffices to show that $\varphi \in X_0$. Indeed, the assumption $\mathsf{N}\varphi \in X_0$ implies that $X_0 \vdash \varphi$ by the Truth axiom and the Modus Ponens inference rule. Therefore, $\varphi \in X_0$ because set $X_0$ is maximal. ⊣

DEFINITION 13.2. For any worlds $w, u \in W$, let $w \sim_a u$ if $\varphi \in u$ for each formula $\mathsf{K}_a\varphi \in w$.

Alternatively, one can define $w \sim_a u$ if sets $w$ and $u$ have the same $\mathsf{K}_a$-formulae. Our approach results in shorter proofs, but it requires to prove the following lemma.

LEMMA 13.2. *The relation $\sim_a$ is an equivalence relation on the set $W$.*

PROOF. *Reflexivity*: Consider any formula $\varphi \in \Phi$. Suppose that $\mathsf{K}_a\varphi \in w$. By Definition 13.2, it suffices to show that $\varphi \in w$. Indeed, assumption $\mathsf{K}_a\varphi \in w$ 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 $\mathsf{K}_a\varphi \in u$. By Definition 13.2, it suffices to show $\varphi \in w$. Suppose the opposite. Then, $\varphi \notin w$. Hence, $w \nvdash \varphi$ because set $w$ is maximal. Thus, $w \nvdash \mathsf{K}_a\varphi$ by the contraposition of the Truth axiom. Then, $\neg\mathsf{K}_a\varphi \in w$ because set $w$ is maximal. Thus, $w \vdash \mathsf{K}_a\neg\mathsf{K}_a\varphi$ by the Negative Introspection axiom and the Modus Ponens inference rule. Hence, $\mathsf{K}_a\neg\mathsf{K}_a\varphi \in w$ because set $w$ is maximal. Then, $\neg\mathsf{K}_a\varphi \in u$ by assumption $w \sim_a u$ and Definition 13.2. Therefore, $\mathsf{K}_a\varphi \notin u$ because set $w$ is consistent, which contradicts the assumption $\mathsf{K}_a\varphi \in u$.

*Transitivity*: Consider any epistemic worlds $w, u, v \in W$ such that $w \sim_a u$ and $u \sim_a v$ and any formula $\mathsf{K}_a\varphi \in w$. By Definition 13.2, it suffices to show $\varphi \in v$. Assumption $\mathsf{K}_a\varphi \in w$ implies $w \vdash \mathsf{K}_a\mathsf{K}_a\varphi$ by Lemma 9.5 and the Modus Ponens inference rule. Thus, $\mathsf{K}_a\mathsf{K}_a\varphi \in w$ because set $w$ is maximal. Hence, $\mathsf{K}_a\varphi \in u$ by the assumption $w \sim_a u$ and Definition 13.2. Therefore, $\varphi \in v$ by the assumption $u \sim_a v$ and Definition 13.2. ⊣

The next step in specifying the canonical model is to define preference relation $\prec_a$ for each agent $a \in \mathcal{A}$, which we do in Definition 13.5. Towards this definition, we first introduce the "emotional base" $\Delta_a$ for each agent $a$. The set $\Delta_a$ contains a formula $\delta$ if agent $a$ could either be possibly happy about $\delta$ or possibly sad about $\neg\delta$.

DEFINITION 13.3. $\Delta_a = \{\delta \in \Phi \mid \overline{\mathsf{N}}\mathsf{H}_a\delta \in X_0\} \cup \{\delta \in \Phi \mid \overline{\mathsf{N}}\mathsf{S}_a\neg\delta \in X_0\}$.

The next lemma holds because set $\Phi$ is countable.

LEMMA 13.3. *The set $\Delta_a$ is countable for each agent $a \in \mathcal{A}$.* $\dashv$

Next, we introduce a total pre-order $\sqsubseteq_a$ on the emotional base $\Delta_a$ of each agent $a \in \mathcal{A}$. Note that this pre-order is different from the canonical preference relation $\prec_a$ that we introduce in Definition 13.5.

DEFINITION 13.4. For any agent $a \in \mathcal{A}$ and any two formulae $\delta, \delta' \in \Delta_a$, let $\delta \sqsubseteq \delta'$ if $\mathsf{N}(\delta \to \delta') \in X_0$.

LEMMA 13.4. $\mathsf{N}(\neg\delta' \to \neg\delta) \in X_0$ *for any agent $a \in \mathcal{A}$ and any two formulae $\delta, \delta' \in \Delta_a$ such that $\delta \sqsubseteq \delta'$.*

PROOF. Formula $(\delta \to \delta') \to (\neg\delta' \to \neg\delta)$ is a propositional tautology. Thus, $\vdash \mathsf{N}((\delta \to \delta') \to (\neg\delta' \to \neg\delta))$ by the Necessitation inference rule. Hence,

$$\vdash \mathsf{N}(\delta \to \delta') \to \mathsf{N}(\neg\delta' \to \neg\delta) \tag{13.1}$$

by the Distributivity axiom and Modus Ponens.

Suppose that $\delta \sqsubseteq \delta'$. Thus, $\mathsf{N}(\delta \to \delta') \in X_0$ by Definition 13.4. Hence, $X_0 \vdash \mathsf{N}(\neg\delta' \to \neg\delta)$ by statement (13.1) and the Modus Ponens inference rule. Therefore, $\mathsf{N}(\neg\delta' \to \neg\delta) \in X_0$ because set $X_0$ is maximal. $\dashv$

LEMMA 13.5. *For any agent $a \in \mathcal{A}$, the relation $\sqsubseteq$ is a total pre-order on the set $\Delta_a$.*

PROOF. We need to show that relation $\sqsubseteq$ is reflexive, transitive, and total.

*Reflexivity*: Consider an arbitrary formula $\delta \in \Phi$. By Definition 13.4, it suffices to show that $\mathsf{N}(\delta \to \delta) \in X_0$. Indeed, the formula $\delta \to \delta$ is a propositional tautology. Thus, $\vdash \mathsf{N}(\delta \to \delta)$ by the Necessitation inference rule. Therefore, $\mathsf{N}(\delta \to \delta) \in X_0$ because set $X_0$ is maximal.

*Transitivity*: Consider arbitrary $\delta_1, \delta_2, \delta_3 \in \Phi$ such that $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ and $\mathsf{N}(\delta_2 \to \delta_3) \in X_0$. By Definition 13.4, it suffices to show that $\mathsf{N}(\delta_1 \to \delta_3) \in X_0$. Indeed, note that the formula

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

is a propositional tautology. Thus,

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

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

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

Then,

$$X_0 \vdash \mathsf{N}((\delta_2 \to \delta_3) \to (\delta_1 \to \delta_3))$$

by the assumption $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ and the Modus Ponens inference rule. Thus,

$$X_0 \vdash \mathsf{N}(\delta_2 \to \delta_3) \to \mathsf{N}(\delta_1 \to \delta_3)$$

by the Distributivity axiom and the Modus Ponens inference rule. Hence, by the assumption $\mathsf{N}(\delta_2 \to \delta_3) \in X_0$ and the Modus Ponens rule,

$$X_0 \vdash \mathsf{N}(\delta_1 \to \delta_3).$$

Therefore, $\mathsf{N}(\delta_1 \to \delta_3) \in X_0$ because set $X_0$ is maximal.

*Totality*: Consider arbitrary formulae $\delta_1, \delta_2 \in \Delta_a$. By Definition 13.4, it suffices to show that either $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ or $\mathsf{N}(\delta_2 \to \delta_1) \in X_0$. By Definition 13.3, without loss of generality, we can assume that one the following three cases take place:

*Case I*: $\overline{\mathsf{N}}\mathsf{H}_a\delta_1, \overline{\mathsf{N}}\mathsf{H}_a\delta_2 \in X_0$. Then, $X_0 \vdash \mathsf{N}(\delta_1 \to \delta_2) \vee \mathsf{N}(\delta_2 \to \delta_1)$ by the first Coherence of Possible Emotions axiom and propositional reasoning. Therefore, because set $X_0$ is maximal, either $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ or $\mathsf{N}(\delta_2 \to \delta_1) \in X_0$.

*Case II*: $\overline{\mathsf{N}}\mathsf{S}_a\neg\delta_1, \overline{\mathsf{N}}\mathsf{S}_a\neg\delta_2 \in X_0$. Similarly to the previous case, we can show that either $\mathsf{N}(\neg\delta_1 \to \neg\delta_2) \in X_0$ or $\mathsf{N}(\neg\delta_2 \to \neg\delta_1) \in X_0$. Without loss of generality, suppose that $\mathsf{N}(\neg\delta_1 \to \neg\delta_2) \in X_0$. Since $(\neg\delta_1 \to \neg\delta_2) \to (\delta_2 \to \delta_1)$ is a propositional tautology, by the Necessitation rule, we have $\vdash \mathsf{N}((\neg\delta_1 \to \neg\delta_2) \to (\delta_2 \to \delta_1))$. Hence, $\vdash \mathsf{N}(\neg\delta_1 \to \neg\delta_2) \to \mathsf{N}(\delta_2 \to \delta_1)$, by the Distributivity axiom and Modus Ponens. Thus, $X_0 \vdash \mathsf{N}(\delta_2 \to \delta_1)$, by the assumption $\mathsf{N}(\neg\delta_1 \to \neg\delta_2) \in X_0$. Therefore, $\mathsf{N}(\delta_2 \to \delta_1) \in X_0$ because set $X_0$ is maximal.

*Case III*: $\overline{\mathsf{N}}\mathsf{H}_a\delta_1, \overline{\mathsf{N}}\mathsf{S}_a\neg\delta_2 \in X_0$. Thus, by the second Coherence of Possible Emotions axiom and propositional reasoning, $X_0 \vdash \mathsf{N}(\delta_1 \to \neg\neg\delta_2) \vee \mathsf{N}(\neg\neg\delta_2 \to \delta_1)$. Hence, either $\mathsf{N}(\delta_1 \to \neg\neg\delta_2) \in X_0$ or $\mathsf{N}(\neg\neg\delta_2 \to \delta_1) \in X_0$ because set $X_0$ is consistent. Then using an argument similar to the one in Case II and propositional tautologies

$$(\delta_1 \to \neg\neg\delta_2) \to (\delta_1 \to \delta_2) \text{ and } (\neg\neg\delta_2 \to \delta_1) \to (\delta_2 \to \delta_1)$$

one can conclude that either $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ or $\mathsf{N}(\delta_2 \to \delta_1) \in X_0$. $\dashv$

We now are ready to define preference relation $\prec_a$ on epistemic worlds of the canonical model.

Definition 13.5. $w \prec_a u$ if there is a formula $\delta \in \Delta_a$ such that $\delta \notin w$ and $\delta \in u$.

Note that the transitivity of the relation $\prec_a$ is not obvious. We prove it as a part of the next lemma.

Lemma 13.6. *$\prec_a$ is a strict partial order on $W$.*

Proof. *Irreflexivity*: Suppose that $w \prec_a w$ for some world $w \in W$. Thus, by Definition 13.5, there exists a formula $\delta \in \Delta_a$ such that $\delta \notin w$ and $\delta \in w$, which is a contradiction.

*Transitivity*: Consider any worlds $w, u, v \in W$ such that $w \prec_a u$ and $u \prec_a v$. It suffices to prove that $w \prec_a v$. Indeed, by Definition 13.5, assumptions $w \prec_a u$ and $u \prec_a v$ imply that there are formulae $\delta_1, \delta_2$ in $\Delta_a$ such that

$$\delta_1 \notin w, \quad \delta_1 \in u, \quad \delta_2 \notin u, \quad \text{and} \quad \delta_2 \in v. \tag{13.2}$$

By Lemma 13.5, either $\delta_1 \sqsubseteq \delta_2$ or $\delta_2 \sqsubseteq \delta_1$. We consider these two cases separately.

*Case I*: $\delta_1 \sqsubseteq \delta_2$. Then, $\mathsf{N}(\delta_1 \to \delta_2) \in X_0$ by Definition 13.5. Hence, $\delta_1 \to \delta_2 \in u$ by Definition 13.1. Thus, $u \vdash \delta_2$ by the part $\delta_1 \in u$ of statement (13.2) and Modus Ponens. Therefore, $\delta_2 \in u$ because set $u$ is maximal, which contradicts the part $\delta_2 \notin u$ of statement (13.2).

*Case II*: $\delta_2 \sqsubseteq \delta_1$. Then, $\mathsf{N}(\delta_2 \to \delta_1) \in X_0$ by Definition 13.5. Thus, $\delta_2 \to \delta_1 \in v$ by Definition 13.1. Hence, $v \vdash \delta_1$ by the part $\delta_2 \in v$ of statement (13.2) and Modus Ponens. Then, $\delta_1 \in v$ because set $v$ is maximal. Therefore, $w \prec_a v$ by Definition 13.5 and the part $\delta_1 \notin w$ of statement (13.2). $\dashv$

Definition 13.6. $\pi(p) = \{w \in W \mid p \in w\}$.

This concludes the definition of the canonical epistemic model with preferences $M(X_0) = (W, \{\sim_a\}_{a \in \mathcal{A}}, \{\prec_a\}_{a \in \mathcal{A}}, \pi)$.

## 14. Properties of a Canonical Model

As usual, the proof of the completeness is centered around an "induction" or "truth" lemma. In our case, this is Lemma 14.11. We precede this lemma with several auxiliary lemmas that are used in the induction step of the proof of Lemma 14.11. For the benefit of the reader, we group these auxiliary lemmas into several subsections. Throughout this section up to and including Lemma 14.11, we assume a fixed canonical model $M(X_0)$.

### 14.1. Properties of Modality $\mathsf{N}$

LEMMA 14.1. *For any worlds $w, u \in W$ and any formula $\varphi \in \Phi$, if $\mathsf{N}\varphi \in w$, then $\varphi \in u$.*

PROOF. Suppose that $\varphi \notin u$. Thus, $u \nvdash \varphi$ because set $u$ is maximal. Hence, $\mathsf{N}\varphi \notin u$ by the Truth axiom. Then, $\mathsf{NN}\varphi \notin X_0$ by Definition 13.1. Thus, $X_0 \nvdash \mathsf{NN}\varphi$ because set $X_0$ is maximal. Hence, $\mathsf{N}\varphi \notin X_0$ by Lemma 9.5. Then, $\neg\mathsf{N}\varphi \in X_0$ because set $X_0$ is maximal. Thus, $X_0 \vdash \mathsf{N}\neg\mathsf{N}\varphi$ by the Negative Introspection axiom and the Modus Ponens inference rule. Hence, $\mathsf{N}\neg\mathsf{N}\varphi \in X_0$ because set $X_0$ is maximal. Then, $\neg\mathsf{N}\varphi \in w$ by Definition 13.1. Therefore, $\mathsf{N}\varphi \notin w$ because set $w$ is consistent. ⊣

LEMMA 14.2. *For any world $w \in W$ and any formula $\varphi \in \Phi$, if $\mathsf{N}\varphi \notin w$, then there is a world $u \in W$ such that $\varphi \notin u$.*

PROOF. Consider the set $X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}$. We start by showing that set $X$ is consistent. Suppose the opposite. Then, there are formulae $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in X_0$ such that $\psi_1, \dots, \psi_n \vdash \varphi$. Thus,

$$\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}\varphi$$

by Lemma 9.4. Hence, $X_0 \vdash \mathsf{N}\varphi$ because $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in X_0$. Then, $X_0 \vdash \mathsf{NN}\varphi$ by Lemma 9.5 and the Modus Ponens inference rule. Thus, $\mathsf{NN}\varphi \in X_0$ because set $X_0$ is maximal. Hence, $\mathsf{N}\varphi \in w$ by Definition 13.1 which contradicts the assumption $\mathsf{N}\varphi \notin w$ of the lemma. Therefore, set $X$ is consistent.

Let set $u$ be a maximum consistent extension of set $X$. Such a set exists by Lemma 9.8. Note that $u \in W$ by Definition 13.1 and the choice of sets $X$ and $u$. Also, $\neg\varphi \in X \subseteq u$ by the choice of sets $X$ and $u$. Therefore, $\varphi \notin u$ because set $u$ is consistent. ⊣

### 14.2. Properties of Modality $\mathsf{K}$

LEMMA 14.3. *For any agent $a \in \mathcal{A}$, any worlds $w, u \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{K}_a\varphi \in w$ and $w \sim_a u$, then $\varphi \in u$.*

PROOF. Assumptions $\mathsf{K}_a\varphi \in w$ and $w \sim_a u$ imply $\varphi \in u$ by Definition 13.2. ⊣

LEMMA 14.4. *For any agent $a \in \mathcal{A}$, any world $w \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{K}_a\varphi \notin w$, then there is a world $u \in W$ such that $w \sim_a u$ and $\varphi \notin u$.*

PROOF. First, we show that the following set of formulae is consistent

$$X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{K}_a\psi \in w\} \cup \{\chi \mid \mathsf{N}\chi \in X_0\}. \tag{14.1}$$

Assume the opposite. Then, there are formulae

$$\mathsf{K}_a\psi_1, \dots, \mathsf{K}_a\psi_m \in w \tag{14.2}$$

and formulae

$$\mathsf{N}\chi_1, \dots, \mathsf{N}\chi_n \in X_0 \tag{14.3}$$

such that

$$\psi_1, \dots, \psi_m, \chi_1, \dots, \chi_n \vdash \varphi.$$

Thus, by Lemma 9.4,

$$\mathsf{K}_a\psi_1, \dots, \mathsf{K}_a\psi_m, \mathsf{K}_a\chi_1, \dots, \mathsf{K}_a\chi_n \vdash \mathsf{K}_a\varphi.$$

Hence, by assumption (14.2),

$$w, \mathsf{K}_a\chi_1, \dots, \mathsf{K}_a\chi_n \vdash \mathsf{K}_a\varphi. \tag{14.4}$$

Consider any integer $i \leq n$. Note that $\mathsf{N}\chi_i \to \mathsf{K}_a\chi_i$ is an instance of the Knowledge of Necessity axiom. Then, $\vdash \mathsf{N}(\mathsf{N}\chi_i \to \mathsf{K}_a\chi_i)$ by the Necessitation inference rule. Thus, $\vdash \mathsf{NN}\chi_i \to \mathsf{NK}_a\chi_i$ by the Distributivity axiom and the Modus Ponens inference rule. Note that $\vdash \mathsf{N}\varphi \to \mathsf{NN}\varphi$ by Lemma 9.5. Hence, $\vdash \mathsf{N}\chi_i \to \mathsf{NK}_a\chi_i$ by the laws of propositional reasoning. Then, $X_0 \vdash \mathsf{NK}_a\chi$ by assumption (14.3). Thus, $\mathsf{NK}_a\chi \in X_0$ because set $X_0$ is maximal. Hence, $\mathsf{K}_a\chi_i \in w$ by Definition 13.1 for any integer $i \leq n$. Then, statement (14.4) implies that $w \vdash \mathsf{K}_a\varphi$. Thus, $\mathsf{K}_a\varphi \in w$ because set $w$ is maximal, which contradicts assumption $\mathsf{K}_a\varphi \notin w$ of the lemma. Therefore, set $X$ is consistent.

By Lemma 9.8, set $X$ can be extended to a maximal consistent set $u$. Then, $\{\chi \mid \mathsf{N}\chi \in X_0\} \subseteq X \subseteq u$ by equation (14.1). Thus, $u \in W$ by Definition 13.1. Also, $\{\psi \mid \mathsf{K}_a\psi \in w\} \subseteq X \subseteq u$ by equation (14.1). Hence, $w \sim_a u$ by Definition 13.2. Finally, $\neg\varphi \in X \subseteq u$ also by equation (14.1). Therefore, $\varphi \notin u$ because set $u$ is consistent. $\dashv$

### 14.3. Common Properties of Modalities $\mathsf{H}$ and $\mathsf{S}$

Recall that by $\mathsf{E}$ we denote one of the two emotional modalities: $\mathsf{H}$ and $\mathsf{S}$.

Lemma 14.5. *For any agent $a \in \mathcal{A}$, any worlds $w, u \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{E}_a\varphi \in w$ and $w \sim_a u$, then $\varphi \in u$.*

Proof. By Lemma 9.2 and the Modus Ponens inference rule, the assumption $\mathsf{E}_a\varphi \in w$ implies $w \vdash \mathsf{K}_a\varphi$. Thus, $\mathsf{K}_a\varphi \in w$ because set $w$ is maximal. Therefore, $\varphi \in u$ by Lemma 14.3 and the assumption $w \sim_a u$. $\dashv$

Lemma 14.6. *For any world $w \in W$, and any formula $\mathsf{E}_a\varphi \in w$, there is a world $u \in W$ such that $\varphi \notin u$.*

Proof. By the Counterfactual axiom and the Modus Ponens inference rule, assumption $\mathsf{E}_a\varphi \in w$ implies $w \vdash \neg\mathsf{N}\varphi$. Thus, $\mathsf{N}\varphi \notin w$ because set $w$ is consistent. Therefore, by Lemma 14.2, there is a world $u \in W$ such that $\varphi \notin u$. $\dashv$

### 14.4. Properties of Modality $\mathsf{H}$

Lemma 14.7. *For any agent $a \in \mathcal{A}$, any worlds $w, u, u' \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{H}_a\varphi \in w$, $\varphi \notin u$ and $\varphi \in u'$, then $u \prec_a u'$.*

Proof. The assumption $\mathsf{H}_a\varphi \in w$ implies $\neg\mathsf{H}_a\varphi \notin w$ because set $w$ is consistent. Thus, $\mathsf{N}\neg\mathsf{H}_a\varphi \notin X_0$ by Definition 13.1 because $w \in W$. Hence, $\neg\mathsf{N}\neg\mathsf{H}_a\varphi \in X_0$ because set $X_0$ is maximal. Then, $\overline{\mathsf{N}}\mathsf{H}_a\varphi \in X_0$ by the definition of modality $\overline{\mathsf{N}}$. Thus, $\varphi \in \Delta_a$ by Definition 13.3. Therefore, $u \prec_a u'$ by Definition 13.5 and the assumptions $\varphi \notin u$ and $\varphi \in u'$ of the lemma. $\dashv$

Lemma 14.8. *For any agent $a \in \mathcal{A}$, any world $w \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{H}_a\varphi \notin w$, $\mathsf{K}_a\varphi \in w$, and $\mathsf{N}\varphi \notin w$, then there are worlds $u, u' \in W$ such that $\varphi \notin u$, $\varphi \in u'$, and $u \not\prec_a u'$.*

Proof. By Lemma 13.5, relation $\sqsubseteq$ forms a total pre-order on set $\Delta_a$. By Lemma 13.3, set $\Delta_a$ is countable. Thus, by the axiom of countable choice, there is an ordering of all formulae in set $\Delta_a$ that agrees with pre-order $\sqsubseteq$. Generally speaking, such an ordering is not unique. We fix any such ordering:

$$\delta_0 \sqsubseteq \delta_1 \sqsubseteq \delta_2 \sqsubseteq \delta_3 \sqsubseteq \dots \tag{14.5}$$

If set $\Delta_a$ is finite, the above ordering has some finite number $n$ of elements. In this case, the ordering is isomorphic to ordinal $n$. Otherwise, it is isomorphic to ordinal $\omega$. Let $\alpha$ be the ordinal which is the type of ordering (14.5). Ordinal $\alpha$ is either finite or is equal to $\omega$.

For any ordinal $k \leq \alpha$, we consider set

$$Y_k = \{\varphi\} \cup \{\neg\delta_i \mid i < k\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}. \tag{14.6}$$

*Claim* 1. *If there is no finite ordinal $k < \alpha$ such that $Y_k$ is consistent and $Y_{k+1}$ is inconsistent, then $Y_\alpha$ is consistent.*

*Proof of Claim.* To prove that $Y_\alpha$ is consistent, it suffices to show that $Y_k$ is consistent for each ordinal $k \leq \alpha$. We prove this by transfinite induction.

**Zero Case**: Suppose that $Y_0$ is not consistent. Thus, there are formulae $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in X_0$ such that

$$\psi_1, \dots, \psi_n \vdash \neg\varphi.$$

Hence,

$$\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}\neg\varphi$$

by Lemma 9.4. Then, $X_0 \vdash \mathsf{N}\neg\varphi$ by the assumption $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in X_0$. Thus, because set $X_0$ is maximal,

$$\mathsf{N}\neg\varphi \in X_0. \tag{14.7}$$

Hence, $\neg\varphi \in w$, by Definition 13.1. Then, $w \vdash \neg\mathsf{K}_a\varphi$ by the contraposition of the Truth axiom and propositional reasoning. Therefore, $\mathsf{K}_a\varphi \notin w$ because set $w$ is consistent, which contradicts the assumption $\mathsf{K}_a\varphi \in w$ of lemma.

**Successor Case**: Suppose that set $Y_k$ is consistent for some $k < \alpha$. By the assumption of the claim, there is no finite ordinal $k < \alpha$ such that $Y_k$ is consistent and $Y_{k+1}$ is inconsistent. Therefore, set $Y_{k+1}$ is consistent.

**Limit Case**: Suppose that set $Y_\omega$ is not consistent. Thus, there are formulae $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in X_0$ and some finite $k < \omega$ such that

$$\psi_1, \dots, \psi_n, \neg\delta_0, \dots, \neg\delta_{k-1} \vdash \neg\varphi.$$

Therefore, set $Y_k$ is not consistent.

$\diamond$

Let $k'$ be any finite ordinal $k' < \alpha$ such that $Y_{k'}$ is consistent and $Y_{k'+1}$ is inconsistent. If such a finite ordinal does not exist, then let $k'$ be ordinal $\alpha$. Note that in either case, set $Y_{k'}$ is consistent by Claim 1. Let $u'$ be any maximal consistent extension of set $Y_{k'}$. Such an extension exists by Lemma 9.8. Note that $\varphi \in Y_{k'} \subseteq u'$ by (14.6) and the choice of $u'$.

*Claim* 2. $u' \in W$.

*Proof of Claim.* Consider any formula $\mathsf{N}\psi \in X_0$. By Definition 13.1, it suffices to show that $\psi \in u'$. Indeed, $\psi \in Y_{k'}$ by (14.6) and the assumption $\mathsf{N}\psi \in X_0$. So $\psi \in u'$ because $Y_{k'} \subseteq u'$ by the choice of $u'$.

$\diamond$

Consider the following set of formulae:

$$Z = \{\neg\varphi\} \cup \{\delta_i \mid k' \leq i < \alpha\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}. \tag{14.8}$$

*Claim* 3. *The set $Z$ is consistent.*

*Proof of Claim.* We consider the following two cases separately:
*Case 1: $k' < \alpha$.* Suppose that set $Z$ is not consistent. Thus, there are finite ordinals $m < \alpha$ and $n < \omega$ and formulae

$$\mathsf{N}\psi_1, \dots \mathsf{N}\psi_n \in X_0 \tag{14.9}$$

such that $k' \leq m$ and

$$\delta_{k'}, \delta_{k'+1}, \dots, \delta_m, \psi_1, \dots, \psi_n \vdash \varphi.$$

Hence, by the Modus Ponens inference rule applied $m - k'$ times,

$$\delta_{k'}, \delta_{k'} \to \delta_{k'+1}, \delta_{k'+1} \to \delta_{k'+2}, \delta_{k'+2} \to \delta_{k'+3}, \dots, \delta_{m-1} \to \delta_m, \\ \psi_1, \dots, \psi_n \vdash \varphi.$$

Then, by Lemma 9.3,

$$\delta_{k'} \to \delta_{k'+1}, \delta_{k'+1} \to \delta_{k'+2}, \dots, \delta_{m-1} \to \delta_m, \psi_1, \dots, \psi_n \vdash \delta_{k'} \to \varphi.$$

Thus, by Lemma 9.4,

$$\mathsf{N}(\delta_{k'} \to \delta_{k'+1}), \mathsf{N}(\delta_{k'+1} \to \delta_{k'+2}), \dots, \mathsf{N}(\delta_{m-1} \to \delta_m),$$
$$\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}(\delta_{k'} \to \varphi).$$

Recall that $\delta_{k'} \sqsubseteq \delta_{k'+1} \sqsubseteq \cdots \sqsubseteq \delta_m$ by assumption (14.5). Hence, it follows that $\mathsf{N}(\delta_{k'} \to \delta_{k'+1}), \mathsf{N}(\delta_{k'+1} \to \delta_{k'+2}), \dots, \mathsf{N}(\delta_{m-1} \to \delta_m) \in X_0$ by Definition 13.4. Then,

$$X_0, \mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}(\delta_{k'} \to \varphi).$$

Thus, by assumption (14.9),

$$X_0 \vdash \mathsf{N}(\delta_{k'} \to \varphi). \tag{14.10}$$

At the same time, $k' < \alpha$ by the assumption of the case. Hence, set $Y_{k'+1}$ is not consistent by the choice of the finite ordinal $k'$. Then, by equation (14.6), there must exist formulae

$$\mathsf{N}\psi'_1, \dots, \mathsf{N}\psi'_p \in X_0 \tag{14.11}$$

such that

$$\neg\delta_0, \neg\delta_1, \neg\delta_2, \dots, \neg\delta_{k'}, \psi'_1, \dots, \psi'_p \vdash \neg\varphi.$$

In other words,

$$\neg\delta_{k'}, \neg\delta_{k'-1}, \neg\delta_{k'-2}, \dots, \neg\delta_0, \psi'_1, \dots, \psi'_p \vdash \neg\varphi.$$

Thus, by applying the Modus Ponens inference rule $k'$ times,

$$\neg\delta_{k'}, \neg\delta_{k'} \to \neg\delta_{k'-1}, \neg\delta_{k'-1} \to \neg\delta_{k'-2}, \dots, \neg\delta_1 \to \neg\delta_0, \psi'_1, \dots, \psi'_p \vdash \neg\varphi.$$

Hence, by Lemma 9.3,

$$\neg\delta_{k'} \to \neg\delta_{k'-1}, \neg\delta_{k'-1} \to \neg\delta_{k'-2}, \dots, \neg\delta_1 \to \neg\delta_0, \psi'_1, \dots, \psi'_n \vdash$$
$$\neg\delta_{k'} \to \neg\varphi.$$

Then, by Lemma 9.4,

$$\mathsf{N}(\neg\delta_{k'} \to \neg\delta_{k'-1}), \mathsf{N}(\neg\delta_{k'-1} \to \neg\delta_{k'-2}), \dots, \mathsf{N}(\neg\delta_1 \to \neg\delta_0),$$
$$\mathsf{N}\psi'_1, \dots, \mathsf{N}\psi'_n \vdash \mathsf{N}(\neg\delta_{k'} \to \neg\varphi).$$

Recall that $\delta_0 \sqsubseteq \delta_1 \sqsubseteq \cdots \sqsubseteq \delta_{k'}$, by (14.5). Thus, it follows that

$$\mathsf{N}(\neg\delta_1 \to \neg\delta_0), \mathsf{N}(\neg\delta_2 \to \neg\delta_1), \dots, \mathsf{N}(\neg\delta_{k'} \to \neg\delta_{k'-1}) \in X_0$$

by Lemma 13.4. Hence,

$$X_0, \mathsf{N}\psi'_1, \dots, \mathsf{N}\psi'_n \vdash \mathsf{N}(\neg\delta_{k'} \to \neg\varphi).$$

Then, by assumption (14.11),

$$X_0 \vdash \mathsf{N}(\neg\delta_{k'} \to \neg\varphi). \tag{14.12}$$

Thus, by item 1 of Lemma 9.6 and statement (14.10),

$$X_0 \vdash \mathsf{N}(\delta_{k'} \leftrightarrow \varphi). \tag{14.13}$$

Note that $\delta_{k'} \in \Delta_a$ because (14.5) is an ordering of set $\Delta_a$. Hence, by Definition 13.3, either $\overline{\mathsf{N}}\mathsf{H}_a\delta_{k'} \in X_0$ or $\overline{\mathsf{N}}\mathsf{S}_a\neg\delta_{k'} \in X_0$. Then, by items 1 and 2 of Lemma 9.7 and statement (14.13), either $X_0 \vdash \overline{\mathsf{N}}\mathsf{H}_a\varphi$ or $X_0 \vdash \overline{\mathsf{N}}\mathsf{S}_a\neg\varphi$. Thus, either $X_0 \vdash \mathsf{N}\overline{\mathsf{N}}\mathsf{H}_a\varphi$ or $X_0 \vdash \mathsf{N}\overline{\mathsf{N}}\mathsf{S}_a\neg\varphi$ by the definition of modality $\overline{\mathsf{N}}$, the Negative Introspection axiom, and the Modus Ponens inference rule. Hence, either $\mathsf{N}\overline{\mathsf{N}}\mathsf{H}_a\varphi \in X_0$ or $\mathsf{N}\overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \in X_0$ because set $X_0$ is maximal. Then, either $\overline{\mathsf{N}}\mathsf{H}_a\varphi \in w$ or $\overline{\mathsf{N}}\mathsf{S}_a\neg\varphi \in w$ by Definition 13.1 and assumption $w \in W$ of the lemma. Thus, $w \vdash \mathsf{K}_a\varphi \to \mathsf{H}_a\varphi$ by the first Emotional Predictability axiom and propositional reasoning. Hence, $w \vdash \mathsf{H}_a\varphi$ by assumption $\mathsf{K}_a\varphi$ of the lemma and the Modus Ponens inference rule. Therefore, $\mathsf{H}_a\varphi \in w$ because set $w$ is maximal, which contradicts assumption $\mathsf{H}_a\varphi \notin w$ of the lemma.
*Case 2:* $\alpha \leq k'$. Recall that $k' \leq \alpha$ by the choice of ordinal $k'$ made after the end of the proof of Claim 1. Thus, $k' = \alpha$. Hence, $Z = \{\neg\varphi\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}$ by equation (14.8). Then, inconsistency of set $Z$ implies that there are formulae

$$\mathsf{N}\psi_1, \dots \mathsf{N}\psi_n \in X_0 \tag{14.14}$$

such that $\psi_1, \dots, \psi_n \vdash \varphi$. Thus, $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}\varphi$ by Lemma 9.4. Hence,

$$X_0 \vdash \mathsf{N}\varphi \tag{14.15}$$

by the assumption (14.14). Then, $X_0 \vdash \mathsf{N}\mathsf{N}\varphi$ by Lemma 9.5 and the Modus Ponens inference rule. Hence, $\mathsf{N}\mathsf{N}\varphi \in X_0$ because set $X_0$ is maximal. Therefore, $\mathsf{N}\varphi \in w$, by Definition 13.1, which contradicts the assumption of the lemma. $\diamond$

Let $u$ be any maximal consistent extension of set $Z$. Note that $\neg\varphi \in Z \subseteq u$ by equation (14.8) and the choice of set $u$.

*Claim* 4. $u \in W$.

*Proof of Claim.* Consider any formula $\mathsf{N}\psi \in X_0$. By Definition 13.1, it suffices to show that $\psi \in u$. Indeed, $\psi \in Z$ by equation (14.8) and the assumption $\mathsf{N}\psi \in X_0$. Thus, $\psi \in u$ because $Z \subseteq u$ by the choice of $u$. $\diamond$

*Claim* 5. $u \not\prec_a u'$.

*Proof of Claim.* Suppose that $u \prec_a u'$. Thus, by Definition 13.5, there is a formula $\delta \in \Delta_a$ such that $\delta \notin u$ and $\delta \in u'$. Recall that (14.5) is an ordering of $\Delta_a$. Hence, there must exist an integer $i < \alpha$ such that

$$\delta_i \notin u \text{ and } \delta_i \in u'. \tag{14.16}$$

We consider the following two cases separately:
*Case 1:* $i < k'$. Then, $\neg\delta_i \in Y_{k'} \subseteq u'$ by equation (14.6) and the choice of set $u'$. Thus, $\delta_i \notin u'$ because set $u'$ is consistent, which contradicts to statement (14.16).
*Case 2:* $k' \leq i$. Then, $\delta_i \in Z \subseteq u$ by equation (14.8) and the choice of set $u$, which contradicts statement (14.16). $\diamond$

This concludes the proof of the lemma. $\dashv$

### 14.5. Properties of Modality $\mathsf{S}$

Lemma 14.9. *For any agent $a \in \mathcal{A}$, any worlds $w, u, u' \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{S}_a\varphi \in w$, $\varphi \in u$ and $\varphi \notin u'$, then $u \prec_a u'$.*

Proof. Note that $\varphi \leftrightarrow \neg\neg\varphi$ is a propositional tautology. Thus, $\vdash \mathsf{N}(\varphi \leftrightarrow \neg\neg\varphi)$ by the Necessitation inference rule. Hence, $\vdash \mathsf{S}_a\varphi \to \mathsf{S}_a\neg\neg\varphi$ by the Substitution axiom and the Modus Ponens inference rule. Then, $w \vdash \mathsf{S}_a\neg\neg\varphi$ by the Modus Ponens inference rule and the assumption $\mathsf{S}_a\varphi \in w$ of the lemma. Thus, $\neg\mathsf{S}_a\neg\neg\varphi \notin w$ because set $w$ is consistent. Hence, $\mathsf{N}\neg\mathsf{S}_a\neg\neg\varphi \notin X_0$ by Definition 13.1. Then, $\neg\mathsf{N}\neg\mathsf{S}_a\neg\neg\varphi \in X_0$ because set $X_0$ is maximal. Thus, $\overline{\mathsf{N}}\mathsf{S}_a\neg\neg\varphi \in X_0$ by the definition of modality $\overline{\mathsf{N}}$. Hence, $\neg\varphi \in \Delta_a$ by Definition 13.3. Therefore, $u \prec_a u'$ by Definition 13.5 and the assumptions $\neg\varphi \notin u$ and $\neg\varphi \in u'$ of the lemma. $\dashv$

Lemma 14.10. *For any agent $a \in \mathcal{A}$, any world $w \in W$, and any formula $\varphi \in \Phi$, if $\mathsf{S}_a\varphi \notin w$, $\mathsf{K}_a\varphi \in w$, and $\mathsf{N}\varphi \notin w$, then there are worlds $u, u' \in W$ such that $\varphi \in u$, $\varphi \notin u'$, and $u \not\prec_a u'$.*

Proof. The proof of this lemma is similar to the proof of Lemma 14.8. Here we outline the differences. The choice of ordering (14.5) and of ordinal $\alpha$ remains the same. Sets $Y_k$ for any ordinal $k \leq \alpha$ is now defined as

$$Y_k = \{\neg\varphi\} \cup \{\neg\delta_i \mid i < k\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}. \tag{14.17}$$

This is different from equation (14.6) because set $Y_k$ now includes $\neg\varphi$ instead of $\varphi$.

The statement of Claim 1 remains the same. The proof of this claim is also the same except for the Zero Case. In Zero Case, the proof is similar to the original till equation (14.7). Because set $Y_k$ now contains $\neg\varphi$ instead of $\varphi$, equation (14.7) will now have the form $\mathsf{N}\varphi \in X_0$. Thus, in our case, $X_0 \vdash \mathsf{NN}\varphi$ by Lemma 9.5 and the Modus Ponens inference rule. Hence, $\mathsf{NN}\varphi \in X_0$ because set $X_0$ is maximal. Then, $\mathsf{N}\varphi \in w$ by Definition 13.1, which contradicts the assumption $\mathsf{N}\varphi \notin w$ of the lemma.

The statement and the proof of Claim 2 remain the same. Set $Z$ will now be defined as

$$Z = \{\varphi\} \cup \{\delta_i \mid k' \leq i < \alpha\} \cup \{\psi \mid \mathsf{N}\psi \in X_0\}. \tag{14.18}$$

This is different from equation (14.8) because set $Z$ now includes $\varphi$ instead of $\neg\varphi$.

The statement of Claim 3 remains the same. Case 1 of the proof of this case is similar to the original proof of Claim 3 till formula (14.12), except for $\varphi$ will be used instead of $\neg\varphi$, and $\neg\varphi$ instead of $\varphi$ everywhere in that part of the proof. In particular formula (14.10) will now have the form

$$X_0 \vdash \mathsf{N}(\delta_{k'} \to \neg\varphi). \tag{14.19}$$

and formula (14.12) will now have the form

$$X_0 \vdash \mathsf{N}(\neg\delta_{k'} \to \varphi). \tag{14.20}$$

The argument after formula (14.12) will change as follows. By item 2 of Lemma 9.6 and statements (14.19) and (14.20),

$$X_0 \vdash \mathsf{N}(\delta_{k'} \leftrightarrow \neg\varphi). \tag{14.21}$$

Note that $\delta_{k'} \in \Delta_a$ because (14.5) is an ordering of set $\Delta_a$. Hence, by Definition 13.3, either $\overline{\mathsf{N}}\mathsf{H}_a\delta_{k'} \in X_0$ or $\overline{\mathsf{N}}\mathsf{S}_a\neg\delta_{k'} \in X_0$. Then, by items 1 or 3 of Lemma 9.7 and statement (14.13), either $X_0 \vdash \overline{\mathsf{N}}\mathsf{H}_a\neg\varphi$ or $X_0 \vdash \overline{\mathsf{N}}\mathsf{S}_a\varphi$. Thus, either $X_0 \vdash \mathsf{N}\overline{\mathsf{N}}\mathsf{H}_a\neg\varphi$ or $X_0 \vdash \mathsf{N}\overline{\mathsf{N}}\mathsf{S}_a\varphi$ by the definition of modality $\overline{\mathsf{N}}$, the Negative Introspection axiom, and the Modus Ponens inference rule. Hence, either $\mathsf{N}\overline{\mathsf{N}}\mathsf{H}_a\neg\varphi \in X_0$ or $\mathsf{N}\overline{\mathsf{N}}\mathsf{S}_a\varphi \in X_0$ because set $X_0$ is maximal. Then, either $\overline{\mathsf{N}}\mathsf{H}_a\neg\varphi \in w$ or $\overline{\mathsf{N}}\mathsf{S}_a\varphi \in w$ by Definition 13.1 and assumption $w \in W$ of the lemma. Thus, $w \vdash \mathsf{K}_a\varphi \to \mathsf{S}_a\varphi$ by the second Emotional Predictability axiom and propositional reasoning. Hence, $w \vdash \mathsf{S}_a\varphi$ by assumption $\mathsf{K}_a\varphi$ of the lemma and the Modus Ponens inference rule. Therefore, $\mathsf{S}_a\varphi \in w$ because set $w$ is maximal, which contradicts assumption $\mathsf{S}_a\varphi \notin w$ of the lemma.

The Case 2 of the proof of Claim 3 will be similar to the original proof of Claim 3 till formula (14.15) except that a formula except for $\varphi$ will be used instead of $\neg\varphi$, and $\neg\varphi$ instead of $\varphi$ everywhere in that part of the proof. Statement (14.15) will now have the form $X_0 \vdash \mathsf{N}\neg\varphi$. From this point, the proof will continue as follows. Statement $X_0 \vdash \mathsf{N}\neg\varphi$ implies that $\mathsf{N}\neg\varphi \in X_0$ because set $X_0$ is maximal. Then, $\neg\varphi \in w$ by Definition 13.1. Hence, $w \vdash \neg\mathsf{K}_a\varphi$ by the contraposition of the Truth axiom. Therefore, $\mathsf{K}_a\varphi \notin w$ because set $w$ is consistent, which contradicts the assumption $\mathsf{K}_a\varphi \in w$ of the lemma.

The statements and the proofs of Claim 4 and Claim 5 remain the same as in the original proof. $\dashv$

### 14.6. Final Steps

We are now ready to state and to prove the "induction" or "truth" lemma.

LEMMA 14.11. *$w \Vdash \varphi$ iff $\varphi \in w$.*

PROOF. We prove the lemma by structural induction on a formula $\varphi$. If $\varphi$ is a propositional variable, then the required follows from Definition 13.6 and item 1 of Definition 3.1. If $\varphi$ is a negation or an implication, then the statement of the lemma follows from the induction hypothesis using items 2 and 3 of Definition 3.1 and the maximality and the consistency of the set $w$ in the standard way.

Suppose that $\varphi$ has the form $\mathsf{K}_a\psi$.

($\Leftarrow$) By Lemma 14.3, assumption $\mathsf{K}_a\psi \in w$ implies that $\psi \in u$ for any world $u \in W$ such that $w \sim_a u$. Thus, by the induction hypothesis, $u \Vdash \psi$ for any world $u \in W$ such that $w \sim_a u$. Therefore, $w \Vdash \mathsf{K}_a\psi$ by item 5 of Definition 3.1.

($\Rightarrow$) Assume that $\mathsf{K}_a\psi \notin w$. Thus, by Lemma 14.4, there is a world $u \in W$ such that $w \sim_a u$ and $\psi \notin u$. Hence, $u \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{K}_a\psi$ by item 5 of Definition 3.1.

If $\varphi$ has the form $\mathsf{N}\psi$, then the proof is similar to the case $\mathsf{K}_a\psi$ except that Lemma 14.1 and Lemma 14.2 are used instead of Lemma 14.3 and Lemma 14.4 respectively. Also, item 4 of Definition 3.1 is used instead of item 5.

Assume that $\varphi$ has the form $\mathsf{H}_a\psi$.

($\Leftarrow$) Assume that $\mathsf{H}_a\psi \in w$. To prove that $w \Vdash \mathsf{H}_a\psi$, we verify conditions (a), (b), and (c) from item 6 of Definition 3.1.

(a) The assumption $\mathsf{H}_a\psi \in w$ implies $w \vdash \mathsf{K}_a\psi$ by Lemma 9.2 and the Modus Ponens inference rule. Thus, $\mathsf{K}_a\psi \in w$ because set $w$ is maximal. Hence, by Lemma 14.3, for any world $u \in W$, if $\psi \in u$, then $w \sim_a u$ . Therefore, by the induction hypothesis, $u \Vdash \psi$ for any world $u \in W$ such that $w \sim_a u$.

(b) By Lemma 14.7, assumption $\mathsf{H}_a\psi \in w$ implies that for any worlds $u, u' \in W$, if $\psi \notin u$ and $\psi \in u'$, then $u \prec_a u'$. Thus, by the induction hypothesis, for any worlds $u, u' \in W$, if $u \nVdash \psi$ and $u' \Vdash \psi$, then $u \prec_a u'$.

(c) By the Counterfactual axiom and the Modus Ponens inference rule, the assumption $\mathsf{H}_a\psi \in w$ implies that $w \vdash \neg\mathsf{N}\psi$. Thus, $\mathsf{N}\psi \notin w$ because set $w$ is consistent. Hence, by Lemma 14.2, there is a world $u \in W$ such that $\psi \notin u$. Therefore, $u \nVdash \psi$ by the induction hypothesis.

($\Rightarrow$) Assume that $\mathsf{H}_a\psi \notin w$. We consider the following three cases separately:

*Case I*: $\mathsf{K}_a\psi \notin w$. Thus, by Lemma 14.4, there is a world $u \in W$ such that $w \sim_a u$ and $\psi \notin u$. Hence, $u \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{H}_a\psi$ by item 6(a) of Definition 3.1.

*Case II*: $\mathsf{N}\psi \in w$. Then, $\psi \in u$ for any world $u \in W$ by Lemma 14.1. Thus, by the induction hypothesis, $u \Vdash \psi$ for any world $u \in W$. Therefore, $w \nVdash \mathsf{H}_a\psi$ by item 6(c) of Definition 3.1.

*Case III*: $\mathsf{K}_a\psi \in w$ and $\mathsf{N}\psi \notin w$. Thus, by the assumption $\mathsf{H}_a\psi \notin w$ and Lemma 14.8, there are worlds $u, u' \in W$ such that $\psi \notin u$, $\psi \in u'$, and $u \nprec_a u'$. Hence, $u \nVdash \psi$ and $u' \Vdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{H}_a\psi$ by item 6(b) of Definition 3.1.

If $\varphi$ has the form $\mathsf{S}_a\psi$, then the argument is similar to the one above, except that Lemma 14.9 and Lemma 14.10 are used instead of Lemma 14.7 and Lemma 14.8 respectively. $\dashv$

THEOREM 14.1 (strong completeness). *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 set $X \cup \{\neg\varphi\}$ is consistent. Thus, by Lemma 9.8, there is a maximal consistent set $w$ such that $X \cup \{\neg\varphi\} \subseteq w$. Consider a canonical epistemic model with preferences $M(w)$. By Lemma 13.1, set $w$ is one of the worlds of this model. Then, $w \Vdash \chi$ for any formula $\chi \in X$ and $w \Vdash \neg\varphi$ by Lemma 14.11. Therefore, $w \nVdash \varphi$ by item 2 of Definition 3.1. $\dashv$

## 15. When You No Longer Know It

In Definition 2.1, we have assumed that each agent $a \in \mathcal{A}$ has the same preference relation $\prec_a$ no matter what the current epistemic world is. Since the model is assumed to be commonly known to all agents, this implies that the preference relation of each agent is also known to all agents. Hence, for instance, it is impossible in our setting to model a situation when Sanaz does *not* know that Pavel prefers epistemic worlds when he receives the gifts to those where he does not.

Note that the assumption that the preferences of each agent are common knowledge is very common in the literature. It appears in game theory when games with imperfect information as discussed [15, p.76]. It was made in previous logical works that studied the interplay between knowledge and preferences [10, 19]. The only existing paper that gives a logical formalism for emotions in an imperfect information setting is [11]. Instead of preferences, it is using an atomic proposition $good_a$ for each agent $a$. These propositions "are used to specify those worlds which are good for an agent" [11, p.827]. Because $good_a$ is a part of the model description, all agents in the system have common knowledge of what worlds are good for which agents.

It is possible, however, to slightly modify the setting in the current article to capture situations when the preferences of the agents are not known to other agents (or even to the agents themselves). To do this, we need to parameterise the preference relations by worlds. Informally, $u \prec_a^w v$ means that "in a world $w$ an agent $a$ prefers a world $v$ over a world $u$". The formal definition of the more general class of models is below.

Definition 15.1. A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \{\prec_a^w\}_{a\in\mathcal{A}}^{w\in W}, \pi)$ is called an generalised 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. $\prec_a^w$ is a strict partial order preference relation on set $W$ for each world $w \in$ and each agent $a \in \mathcal{A}$,
4. $\pi(p)$ is a subset of $W$ for each propositional variable $p$.

For this more general class of models, Definition 3.1 can be modified by using relation $u \prec_a^w u'$ instead of relation $u \prec_a u'$ in items 6(b) and 7(b) of the definition.

It is easy to see that in this more general setting the Emotional Introspection, the Coherence of Possible Emotions, the Emotional Predictability axioms are no longer valid. The title of the current section reflects the fact that the Emotional Introspection axiom, which is the title of the whole article, no longer holds.

We do not believe that the existing proof of the completeness can be adopted to show the completeness of the logical system containing the remaining axioms with respect to generalised epistemic model with preferences.

## 16. Conclusion

In this article, we proposed formal semantics for happiness and sadness, proved that these two notions are not definable through each other and gave a complete logical system capturing the properties of these notions. The approach to happiness that we advocated could be captured by the famous saying "Success is getting what you want, happiness is wanting what you get". Although popular, this view is not the only possible one. As we mentioned in the introduction, some view happiness as "getting what you want".

As defined in this article, happiness and sadness are grounded in an agent's knowledge. We think that an interesting next step could be exploring belief-based happiness and sadness. A framework for beliefs, similar to our epistemic models with preferences, has been proposed in [10].

## References

[1] Adam, C., A. Herzig, and D. Longin, "A logical formalization of the occ theory of emotions", *Synthese* 168 (2), 2009: 201–248. DOI: 10.1007/s11229-009-9460-9

[2] Ahmed, S., *The Promise of Happiness*, Duke University Press, 2010.

[3] Argyle, M., *The Psychology of Happiness*, Routledge, 2013.

[4] Bok, S., *Exploring Happiness: From Aristotle to Brain Science*, Yale University Press, 2010.

[5] Bruni, L., and P. L. Porta, *Economics and Happiness: Framing the Analysis*, OUP Oxford, 2005.

[6] Doyle, J., Y. Shoham, and M. P. Wellman, "A logic of relative desire", pages 16–31 in *International Symposium on Methodologies for Intelligent Systems*, Springer, 1991. DOI: 10.1007/3-540-54563-8_65

[7] Feldman, F., "An improved whole life satisfaction theory of happiness?", *International Journal of Wellbeing* 9 (2), 2019. DOI: 10.5502/ijw.v9i2.762

[8] Frey, B. S., *Happiness: A Revolution in Economics*, MIT Press, 2010.

[9] Lang, J., L. Van Der Torre, and E. Weydert, "Utilitarian desires", *Autonomous Agents and Multi-Agent Systems* 5 (3), 2002: 329–363.

[10] Liu, F., *Reasoning about Preference Dynamics*, volume 354 of Springer Science & Business Media, 2011.

[11] Lorini, E., and F. Schwarzentruber, "A logic for reasoning about counterfactual emotions", *Artificial Intelligence* 175 (3), 2011: 814–847. DOI: 10.1016/j.artint.2010.11.022

[12] Mendelson, E., *Introduction to Mathematical Logic*, CRC Press, Boca Raton, Florida, 2009.

[13] Meyer, J.-J. Ch., "Reasoning about emotional agents", pages 129–133 in *Proceedings of the 16th European Conference on Artificial Intelligence*, 2004.

[14] Ortony, A., G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1988.

[15] Osborne, M. J., and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.

[16] Steunebrink, B. R., M. Dastani, and J.-J. Ch. Meyer, "A logic of emotions for intelligent agents", page 142 in *Proceedings of the National Conference on Artificial Intelligence*, volume 22, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, 1999 (2007).

[17] Stewart, M. E., K. P. Ebmeier, and I. J. Deary, "Personality correlates of happiness and sadness: Epq-r and tpq compared", *Personality and Individual Differences* 38 (5), 2005: 1085–1096. DOI: 10.1016/j.paid.2004.07.007

[18] Suikkanen, J., "An improved whole life satisfaction theory of happiness", *International Journal of Wellbeing* 1 (1), 2011. DOI: 10.5502/ijw.v9i2.762

[19] Van Benthem, J., and Liu, F., "Dynamic logic of preference upgrade", *Journal of Applied Non-Classical Logics* 17 (2), 2007: 157–182. DOI: 10.3166/jancl.17.157-182

## Appendices: Proofs of Propositions

### A. The Gift Scenario

PROPOSITION 4.2.
*$z \Vdash \mathsf{H}_s$("Pavel received a gift from Sanaz") iff $z \in \{w\}$.*

PROOF. ($\Rightarrow$) Suppose that $z \notin \{w\}$. Thus, either $z \in \{t\}$ or $z \in \{u, v\}$. We consider these two cases separately.

*Case I*: $z \in \{t\}$. Then, $z \nVdash$ "Pavel received a gift from Sanaz", see Figure 2. Therefore, $z \Vdash \mathsf{H}_s$("Pavel received a gift from Sanaz") by item 6(a) of Definition 3.1 and because $z \sim_s z$.

*Case II*: $z \in \{u, v\}$. Then, $z \sim_s v$, see Figure 2. Note that, see again Figure 2, $v \nVdash$ "Pavel received a gift from Sanaz". Therefore, item 6(a) of Definition 3.1 implies $z \nVdash \mathsf{H}_s$("Pavel received a gift from Sanaz"), .

($\Leftarrow$) To show that $w \Vdash \mathsf{H}_s$("Pavel received a gift from Sanaz"), we verify conditions (a), (b), and (c) of item 6 of Definition 3.1:

*Condition a*: Consider any world $z$ such that $w \sim_s z$. It suffices to show that $z \Vdash$ "Pavel received a gift from Sanaz". Indeed, assumption $w \sim_s z$ implies that $z = w$, see Figure 2. Note that

$$w \Vdash \text{"Pavel received a gift from Sanaz"},$$

see again Figure 2.

*Condition b*: The proof is similar to the proof of *Condition b* in Proposition 4.1.

*Condition c*: $t \nVdash$ "Pavel received a gift from Sanaz". $\dashv$

The next proposition shows that Sanaz is happy that Pavel is happy only if she gets the thank-you card and, thus, she knows that he received the gift.

PROPOSITION 4.3.
*$z \Vdash \mathsf{H}_s\mathsf{H}_p$("Pavel received a gift from Sanaz") iff $z \in \{w\}$.*

PROOF. The statement $x \Vdash$ "Pavel received a gift from Sanaz" is true iff $x \in \{w, u\}$, see Figure 2. Thus, by Proposition 4.1, for any world $x \in W$,

$$x \Vdash \text{"Pavel received a gift from Sanaz" iff } \qquad x \Vdash \mathsf{H}_p(\text{"Pavel received a gift from Sanaz"}).$$

Hence, by Lemma 3.1, for any world $x \in W$,

$$x \Vdash \mathsf{H}_s(\text{“Pavel received a gift from Sanaz”}) \text{ iff } x \Vdash \mathsf{H}_s\mathsf{H}_p(\text{“Pavel received a gift from Sanaz”}).$$

Therefore, $z \Vdash \mathsf{H}_s\mathsf{H}_p$(“Pavel received a gift from Sanaz”) iff $z \in \{w\}$ and Proposition 4.2. ⊣

Proposition 4.4.
*$z \nVdash \mathsf{H}_p\mathsf{H}_s$(“Pavel received a gift from Sanaz”) for each $z \in \{w, u, v, t\}$.*

Proof. We consider the following two cases separately:

*Case I*: $z \in \{w, u\}$. Then, $z \sim_p u$, see Figure 2. By Proposition 4.2,

$$u \nVdash \mathsf{H}_s(\text{“Pavel received a gift from Sanaz”}).$$

Therefore, $z \nVdash \mathsf{H}_p\mathsf{H}_s$(“Pavel received a gift from Sanaz”) by item 6(a) of Definition 3.1 and the statement $z \sim_p u$.

*Case II*: $z \in \{v, t\}$. Then, $z \nVdash \mathsf{H}_s$(“Pavel received a gift from Sanaz”) by Proposition 4.2. Thus, $z \nVdash \mathsf{H}_p\mathsf{H}_s$(“Pavel received a gift from Sanaz”) by item 6(a) of Definition 3.1. ⊣

The proof of the next statement is similar to the proof of Proposition 4.4 except that it refers to Proposition 4.3 instead of Proposition 4.2.

Proposition 4.5. *$z \nVdash \mathsf{H}_p\mathsf{H}_s\mathsf{H}_p$(“Pavel received a gift from Sanaz”) for each epistemic world $z \in \{w, u, v, t\}$.* ⊣

The next proposition states that Sanaz is sad about Pavel not receiving the gift only if she does not send it. Informally, this proposition is true because Sanaz cannot distinguish a world $v$ in which the gift is lost from a world $u$ in which the card is lost.

Proposition 4.6.
*$z \Vdash \mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”) iff $z \in \{t\}$.*

Proof. ($\Rightarrow$) Suppose that $z \notin \{t\}$. Thus, either $z \in \{w\}$ or $z \in \{u, v\}$. We consider the these two cases separately:

*Case I*: $z \in \{w\}$. Then, $z \Vdash$ “Pavel received a gift from Sanaz”, see Figure 2. Thus, $z \nVdash \neg$(“Pavel received a gift from Sanaz”) by item 2 of Definition 3.1. Therefore, $z \nVdash \mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”) by item 7(a) of Definition 3.1.

*Case II*: $z \in \{u, v\}$. Then, $z \sim_s u$, see Figure 2. Note that

$$u \Vdash \text{“Pavel received a gift from Sanaz”},$$

see Figure 2. Thus, $u \nVdash \neg$(“Pavel received a gift from Sanaz”) by item 2 of Definition 3.1. So $z \nVdash \mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”), by item 7(a) of Definition 3.1 and the statement $z \sim_s u$.

($\Leftarrow$) To prove that $t \Vdash \mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”), we verify conditions (a), (b), and (c) of item 7 in Definition 3.1 separately:

*Condition a*: Consider any world $z'$ such that $t \sim_s z'$. It suffices to show that $z' \Vdash \neg$(“Pavel received a gift from Sanaz”). Indeed, note that

$$t \nVdash (\text{“Pavel received a gift from Sanaz”}),$$

see Figure 2. Then, $t \Vdash \neg$(“Pavel received a gift from Sanaz”) by item 2 of Definition 3.1. Also note that the assumption $t \sim_s z'$ implies that $t = z'$, see Figure 2. Therefore, $z' \Vdash \neg$(“Pavel received a gift from Sanaz”).

*Condition b*: Consider any two epistemic worlds $x, y$ such that

$$x \Vdash \neg(\text{“Pavel received a gift from Sanaz”}),$$
$$y \nVdash \neg(\text{“Pavel received a gift from Sanaz”}).$$

To verify the condition, it suffices to show that $x \prec_s y$. Indeed, by item 2 of Definition 3.1,

$$x \nVdash \text{“Pavel received a gift from Sanaz”},$$
$$y \Vdash \text{“Pavel received a gift from Sanaz”}.$$

Thus, $x \in \{t, v\}$ and $y \in \{w, u\}$, see Figure 2. Note that $\{t, v\} \prec_s \{w, u\}$, see also Figure 2. Therefore, $x \prec_s y$.

*Condition c*: Note that $w \Vdash$ “Pavel received a gift from Sanaz”. So,

$$w \nVdash \neg(\text{“Pavel received a gift from Sanaz”})$$

by item 2 of Definition 3.1. $\dashv$

By Proposition 4.6, Sanaz is sad about Pavel not receiving the gift only if she does not send it. Since Pavel cannot distinguish a world $t$ in which the gift is sent from a world $v$ in which it is lost, Pavel cannot know that Sanaz is sad. This is formally captured in the next proposition.

PROPOSITION 4.7. *$z \nVdash \mathsf{K}_p\mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”) for each epistemic world $z \in \{w, u, t, v\}$.*

Proof. We consider the following two cases separately:

*Case I*: $z \in \{w, u\}$. Then, $z \nVdash \mathsf{S}_s\neg$(“Pavel received a gift from Sanaz”) by Proposition 4.6. Therefore,

$$z \nVdash \mathsf{K}_p\mathsf{S}_s\neg(\text{“Pavel received a gift from Sanaz”})$$

by item 5 of Definition 3.1.

*Case II*: $z \in \{t, v\}$. Then, $z \sim_p v$, see Figure 2. Note that

$$v \nVdash \mathsf{S}_s\neg(\text{“Pavel received a gift from Sanaz”})$$

by Proposition 4.6. Therefore,

$$z \nVdash \mathsf{K}_p\mathsf{S}_s\neg(\text{“Pavel received a gift from Sanaz”})$$

by item 5 of Definition 3.1 and the statement $z \sim_p v$. ⊣

Proposition 4.8.
*$z \Vdash \mathsf{S}_p\neg$(“Pavel received a gift from Sanaz”) iff $z \in \{v, t\}$.*

Proof. ($\Rightarrow$) Suppose that $z \notin \{v, t\}$. Thus, $z \in \{w, u\}$. Hence, see Figure 2, $z \Vdash$ “Pavel received a gift from Sanaz”. Then, by item 2 of Definition 3.1,

$$z \nVdash \neg(\text{“Pavel received a gift from Sanaz”}).$$

Therefore, $z \nVdash \mathsf{S}_p\neg$(“Pavel received a gift from Sanaz”), by item 7(a) of Definition 3.1.

($\Leftarrow$) Let $z \in \{v, t\}$. We verify conditions (a), (b), (c) from item 7 of Definition 3.1 to prove that $z \Vdash \mathsf{S}_p\neg$(“Pavel received a gift from Sanaz”):

*Condition a*: Consider any world $z'$ such that $z \sim_p z'$. It suffices to show that $z' \Vdash \neg$(“Pavel received a gift from Sanaz”). Indeed, the assumptions $z \in \{v, t\}$ and $z \sim_p z'$ imply that $z' \in \{v, t\}$, see Figure 2. Thus, $z' \nVdash$ “Pavel received a gift from Sanaz”, see again Figure 2. Therefore, by item 2 of Definition 3.1, $z' \Vdash \neg$(“Pavel received a gift from Sanaz”).

*Condition b*: The proof is similar to the proof of *Condition b* in Proposition 4.6.

*Condition c*: Note that $w \Vdash$ “Pavel received a gift from Sanaz”. So,

$$w \nVdash \neg(\text{“Pavel received a gift from Sanaz”})$$

by item 2 of Definition 3.1. ⊣

PROPOSITION 4.9.
$z \Vdash \mathsf{S}_s\mathsf{S}_p\neg$(*"Pavel received a gift from Sanaz"*) *iff* $z \in \{t\}$.

PROOF. Note that $x \nVdash$ "Pavel received a gift from Sanaz" iff $x \in \{t, v\}$, see Figure 2. Thus, $x \Vdash \neg$("Pavel received a gift from Sanaz") iff $x \in \{t, v\}$ by item 2 of Definition 3.1. Thus, by Proposition 4.8, for any world $x \in W$,

$$x \Vdash \neg(\text{"Pavel received a gift from Sanaz"}) \text{ iff } x \Vdash \mathsf{S}_p\neg(\text{"Pavel received a gift from Sanaz"}).$$

Hence, by Lemma 3.1, for any world $z \in W$,

$$z \Vdash \mathsf{S}_s\neg(\text{"Pavel received a gift from Sanaz"}) \text{ iff } z \Vdash \mathsf{S}_s\mathsf{S}_p\neg(\text{"Pavel received a gift from Sanaz"}).$$

So, by Proposition 4.6, $z \Vdash \mathsf{S}_s\mathsf{S}_p\neg$("Pavel received a gift from Sanaz") iff $z \in \{t\}$. ⊣

PROPOSITION 4.10. $z \nVdash \mathsf{K}_p\mathsf{S}_s\mathsf{S}_p\neg$(*"Pavel received a gift from Sanaz"*) *for each epistemic world* $z \in \{w, u, v, t\}$.

PROOF. We consider the following two cases separately:

*Case I*: $z \in \{w, u\}$. Then,

$$z \nVdash \mathsf{S}_s\mathsf{S}_p\neg(\text{"Pavel received a gift from Sanaz"})$$

by Proposition 4.9. Therefore,

$$z \nVdash \mathsf{K}_p\mathsf{S}_s\mathsf{S}_p\neg(\text{"Pavel received a gift from Sanaz"})$$

by item 5 of Definition 3.1.

*Case II*: $z \in \{t, v\}$. Then, $z \sim_p v$, see Figure 2. Note that, by Proposition 4.9, $v \nVdash \mathsf{S}_s\mathsf{S}_p\neg$("Pavel received a gift from Sanaz"). Therefore, by item 5 of Definition 3.1,

$$z \nVdash \mathsf{K}_p\mathsf{S}_s\mathsf{S}_p\neg(\text{"Pavel received a gift from Sanaz"}).$$ ⊣

## B. The Battle of Cuisines Scenario

PROPOSITION 5.2.
(Russian, Russian) $\nVdash \mathsf{H}_s$(*"Sanaz is in the Russian restaurant"*).

Proof. Note that

$$(\text{Iranian}, \text{Iranian}) \nVdash \text{“Sanaz is in the Russian restaurant”},$$
$$(\text{Russian}, \text{Iranian}) \Vdash \text{“Sanaz is in the Russian restaurant”}.$$

At the same time, see Table 1,

$$u_s(\text{Russian}, \text{Iranian}) = 0 < 1 = u_s(\text{Iranian}, \text{Iranian}).$$

Hence, $(\text{Iranian}, \text{Iranian}) \not\prec_s (\text{Russian}, \text{Iranian})$. Therefore, by item 6(b) of Definition 3.1,

$$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_s(\text{“Sanaz is in the Russian restaurant”}). \quad \dashv$$

Proposition 5.3.
*$(x, y) \Vdash \mathsf{H}_s$(“Sanaz and Pavel are in the same restaurant”) iff $x = y$.*

Proof. ($\Rightarrow$) Suppose that $x \neq y$. Then,

$$(x, y) \nVdash \text{“Sanaz and Pavel are in the same restaurant”},$$
$$(x, x) \Vdash \text{“Sanaz and Pavel are in the same restaurant”}.$$

Also, $u_s(x, y) = 0 < 1 \leq u_s(x, x)$ because $x \neq y$, see Table 1. Hence $(x, x) \not\prec_s (x, y)$. Therefore,

$$(x, y) \nVdash \mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”})$$

by item 6(b) of Definition 3.1. ($\Leftarrow$) Suppose $x = y$. We verify conditions (a), (b), and (c) from item 6 of Definition 3.1 to prove that

$$(x, y) \Vdash \mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}).$$

*Condition a*: Since this is a model with perfect information, it suffices to show that $(x, y) \Vdash$ “Sanaz and Pavel are in the same restaurant”, which is true due to the assumption $x = y$.

*Condition b*: Consider any two worlds $(x_1, y_1), (x_2, y_2) \in W$ such that

$$(x_1, y_1) \nVdash \text{“Sanaz and Pavel are in the same restaurant”}, \tag{B.1}$$
$$(x_2, y_2) \Vdash \text{“Sanaz and Pavel are in the same restaurant”}. \tag{B.2}$$

It suffices to show that $(x_1, y_1) \prec_s (x_2, y_2)$. Indeed, statements (B.1) and (B.2) imply that $x_1 \neq y_1$ and $x_2 = y_2$, respectively. Thus, $u_s(x_1, y_1) = 0 < 1 \leq u_s(x_2, y_2)$, see Table 1. Therefore, $(x_1, y_1) \prec_s (x_2, y_2)$.

*Condition c*: Note that

$$(\text{Russian}, \text{Iranian}) \nVdash \text{“Sanaz and Pavel are in the same restaurant”}. \quad \dashv$$

The proof of the next proposition is similar to the proof of the one above.

PROPOSITION 5.4.
*$(x, y) \Vdash \mathsf{H}_p$("Sanaz and Pavel are in the same restaurant") iff $x = y$.* $\dashv$

PROPOSITION 5.5.
*$(x, y) \Vdash \mathsf{H}_p\mathsf{H}_s$("Sanaz and Pavel are in the same restaurant") iff $x = y$.*

PROOF. Note that

$$(x, y) \Vdash \text{“Sanaz and Pavel are in the same restaurant”}$$

iff $x = y$. Thus, by Proposition 5.3, for any world $(x, y) \in W$,

$$\begin{aligned}&(x, y) \Vdash \text{“Sanaz and Pavel are in the same restaurant” iff}\\ &\qquad (x, y) \Vdash \mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}).\end{aligned}$$

Hence, by Lemma 3.1, for any world $(x, y) \in W$,

$$\begin{aligned}&(x, y) \Vdash \mathsf{H}_p(\text{“Sanaz and Pavel are in the same restaurant”}) \text{ iff}\\ &\qquad (x, y) \Vdash \mathsf{H}_p\mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}).\end{aligned}$$

Thus, $(x, y) \Vdash \mathsf{H}_p\mathsf{H}_s$("Sanaz and Pavel are in the same restaurant") iff $x = y$, by Proposition 5.4. $\dashv$

PROPOSITION 5.6. *$(x, y) \Vdash \mathsf{H}_s\mathsf{H}_p\mathsf{H}_s$("Sanaz and Pavel are in the same restaurant") iff $x = y$.*

PROOF. Note that

$$(x, y) \Vdash \text{“Sanaz and Pavel are in the same restaurant” iff } x = y.$$

Thus, by Proposition 5.5, for any world $(x, y) \in W$,

$$\begin{aligned}&(x, y) \Vdash \text{“Sanaz and Pavel are in the same restaurant” iff}\\ &\qquad (x, y) \Vdash \mathsf{H}_p\mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}).\end{aligned}$$

Hence, by Lemma 3.1, for any world $(x, y) \in W$,

$$\begin{aligned}&(x, y) \Vdash \mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}) \text{ iff}\\ &\qquad (x, y) \Vdash \mathsf{H}_s\mathsf{H}_p\mathsf{H}_s(\text{“Sanaz and Pavel are in the same restaurant”}).\end{aligned}$$

Then, $(x, y) \Vdash \mathsf{H}_s\mathsf{H}_p\mathsf{H}_s$("Sanaz and Pavel are in the same restaurant") iff $x = y$ by Proposition 5.3. $\dashv$

Proposition 5.7.
*$(x, y) \Vdash \mathsf{H}_s$("Sanaz and Pavel are in the Russian restaurant") is true if and only if $x = y =$ Russian.*

Proof. ($\Rightarrow$) By item 6(a) of Definition 3.1, the assumption of the proposition

$$(x, y) \Vdash \mathsf{H}_s(\text{"Sanaz and Pavel are in the Russian restaurant"})$$

implies that

$$(x, y) \Vdash \text{"Sanaz and Pavel are in the Russian restaurant"}.$$

Therefore, $x = y =$ Russian.

($\Leftarrow$) Suppose $x = y =$ Russian. To prove that

$$(x, y) \Vdash \mathsf{H}_s(\text{"Sanaz and Pavel are in the Russian restaurant"}),$$

we verify conditions (a), (b), and (c) from item 6 of Definition 3.1:

*Condition a*: Since this is a model with perfect information, it suffices to show that $(x, y) \Vdash$ "Sanaz and Pavel are in the Russian restaurant", which is true due to the assumption $x = y =$ Russian.

*Condition b*: Consider any two worlds $(x_1, y_1), (x_2, y_2) \in W$ such that

$$(x_1, y_1) \nVdash \text{"Sanaz and Pavel are in the Russian restaurant"}, \tag{B.3}$$

$$(x_2, y_2) \Vdash \text{"Sanaz and Pavel are in the Russian restaurant"}. \tag{B.4}$$

It suffices to show that $(x_1, y_1) \prec_s (x_2, y_2)$. Indeed, statement (B.3) implies that $u_s(x_1, y_1) \leq 1$, see Table 1. Similarly, statement (B.4) implies that $u_s(x_2, y_2) = 3$. Thus, $u_s(x_1, y_1) \leq 1 < 3 = u_s(x_2, y_2)$. Therefore, $(x_1, y_1) \prec_s (x_2, y_2)$.

*Condition c*: Note that

$$(\text{Russian}, \text{Iranian}) \nVdash \text{"Sanaz and Pavel are in the Russian restaurant"}.$$

$\dashv$

Proposition 5.8.

$$(\text{Russian}, \text{Russian}) \nVdash \mathsf{H}_p(\textit{"Sanaz and Pavel are in the Russian restaurant"}).$$

PROOF. Note that $u_p(\text{Iranian}, \text{Iranian}) = 3 > 1 = u_p(\text{Russian}, \text{Russian})$. Thus,

$$(\text{Iranian}, \text{Iranian}) \not\prec_p (\text{Russian}, \text{Russian}).$$

Therefore, the proposition is true by item 6(b) of Definition 3.1. ⊣

PROPOSITION 5.9.
$(\text{Russian}, \text{Russian}) \nVdash$
$\mathsf{H}_p\mathsf{H}_s$(*"Sanaz and Pavel are in the Russian restaurant"*).

PROOF. Note that, by Proposition 5.7:

$$(\text{Iranian}, \text{Iranian}) \nVdash \mathsf{H}_s(\text{"Sanaz and Pavel are in the Russian restaurant"})$$

and

$$(\text{Russian}, \text{Russian}) \Vdash \mathsf{H}_s(\text{"Sanaz and Pavel are in the Russian restaurant"}).$$

At the same time,

$$u_p(\text{Iranian}, \text{Iranian}) = 3 > 1 = u_p(\text{Russian}, \text{Russian}).$$

Thus, $(\text{Iranian}, \text{Iranian}) \not\prec_p (\text{Russian}, \text{Russian})$. Therefore, the proposition is true by item 6(b) of Definition 3.1. ⊣

PROPOSITION 5.10.
$(x, y) \Vdash \mathsf{S}_s$(*"Sanaz and Pavel are in different restaurants"*) *iff* $x \neq y$.

PROOF. (⇒) Suppose that $x = y$. Thus,

$$(x, y) \nVdash \text{"Sanaz and Pavel are in different restaurants"}.$$

Thus, by item 7(a) of Definition 3.1,

$$(x, y) \nVdash \mathsf{S}_s(\text{"Sanaz and Pavel are in different restaurants"}).$$

(⇐) Suppose that $x \neq y$. To prove that

$$(x, y) \Vdash \mathsf{S}_s(\text{"Sanaz and Pavel are in different restaurants"}),$$

we verify conditions (a), (b), and (c) from item 7 of Definition 3.1:
*Condition a*: Since this is a model with perfect information, it suffices to show that $(x, y) \Vdash$ "Sanaz and Pavel are in different restaurants", which is true due to the assumption $x \neq y$.

*Condition b*: The proof of this condition is similar to the proof of *Condition b* in the proof of Proposition 5.3.

*Condition c*:
(Iranian, Iranian) $\nVdash$ "Sanaz and Pavel are in different restaurants". ⊣

The proof of the following statement is similar to the proof of the one above.

PROPOSITION 5.11.
$(x, y) \Vdash \mathsf{S}_p$(*"Sanaz and Pavel are in different restaurants"*) *iff* $x \neq y$.

PROPOSITION 5.12.
$(x, y) \Vdash \mathsf{S}_p\mathsf{S}_s$(*"Sanaz and Pavel are in different restaurants"*) *iff* $x \neq y$.

PROOF. Note that

$$(x, y) \Vdash \text{"Sanaz and Pavel are in different restaurants"}$$

iff $x \neq y$. Thus, by Proposition 5.10, for any world $(x, y) \in W$,

$$(x, y) \Vdash \text{"Sanaz and Pavel are in different restaurants" iff}$$
$$(x, y) \Vdash \mathsf{S}_s(\text{"Sanaz and Pavel are in different restaurants"}).$$

Hence, by Lemma 3.1, for any world $(x, y) \in W$,

$$(x, y) \Vdash \mathsf{S}_p(\text{"Sanaz and Pavel are in different restaurants"}) \text{ iff}$$
$$(x, y) \Vdash \mathsf{S}_p\mathsf{S}_s(\text{"Sanaz and Pavel are in different restaurants"}).$$

Therefore, $(x, y) \Vdash \mathsf{S}_p\mathsf{S}_s$("Sanaz and Pavel are in different restaurants") iff $x = y$ by Proposition 5.11. ⊣

### B.1. The Lottery Scenario

PROPOSITION 6.2. $u \nVdash \mathsf{H}_s$(*"Pavel lost the lottery"*).

PROOF. Note that

$$w \nVdash \text{"Pavel lost the lottery"},$$
$$v \Vdash \text{"Pavel lost the lottery"},$$

and $w \not\sim_s v$, see Figure 4. Therefore, $u \nVdash \mathsf{H}_s$("Pavel lost the lottery") by item 6(b) of Definition 3.1. ⊣

PROPOSITION 6.3. $u \nVdash \mathsf{K}_p\mathsf{H}_s$(*"Sanaz won the lottery"*).

PROOF. Note that $u \sim_p v$, see Figure 4. Also, by Proposition 6.1,

$$v \nVdash \mathsf{H}_s(\text{“Sanaz won the lottery”}).$$

Then, $u \nVdash \mathsf{K}_p\mathsf{H}_s$(“Sanaz won the lottery”) by item 5 of Definition 3.1. ⊣

PROPOSITION 6.4. *$u \Vdash \mathsf{S}_p$(“Pavel lost the lottery”).*

PROOF. It suffices to verify conditions (a), (b), and (c) from item 7 of Definition 3.1:

*Condition a*: Consider any world $y$ such that $u \sim_p y$. We will show that

$$y \Vdash \text{“Pavel lost the lottery”}.$$

Indeed, assumption $u \sim_p y$ implies that $y \in \{u, v\}$, see Figure 4. Therefore, see again Figure 4, $y \Vdash$ “Pavel lost the lottery”.

*Condition b*: Consider any $y, z$ such that $y \Vdash$ “Pavel lost the lottery” and $z \nVdash$ “Pavel lost the lottery”. We will show that $y \prec_p z$. Indeed, the assumption

$$y \Vdash \text{“Pavel lost the lottery”}$$

implies $y \in \{u, v\}$, see Figure 4. The assumption

$$z \nVdash \text{“Pavel lost the lottery”}$$

similarly implies that $z = w$. Statements $y \in \{u, v\}$ and $z = w$ imply that $y \prec_s z$, see again Figure 4.

*Condition c*: $w \nVdash$ “Pavel lost the lottery”. ⊣

PROPOSITION 6.5. *$u \Vdash \mathsf{K}_s\mathsf{S}_p$(“Pavel lost the lottery”).*

PROOF. Consider any world $y$ such that $u \sim_s y$. By item 5 of Definition 3.1, it suffices to show that $y \Vdash \mathsf{S}_p$(“Pavel lost the lottery”). Indeed, assumption $u \sim_s y$ implies that $u = y$, see Figure 4. Therefore, $y \Vdash \mathsf{S}_p$(“Pavel lost the lottery”) by Proposition 6.4. ⊣

SANAZ AZIMIPOUR
Berlin University of Applied Sciences, Germany
`Sanaz.a234@gmail.com`

PAVEL NAUMOV
University of Southampton, United Kingdom
`p.naumov@soton.ac.uk`
https://orcid.org/0000-0003-1687-045X